A VOCABULARY
INFORMATION TEST

BY
ANGELINA LOUISA WEEKS, Ph.D.

To My Sister

Mary Grace Weeks
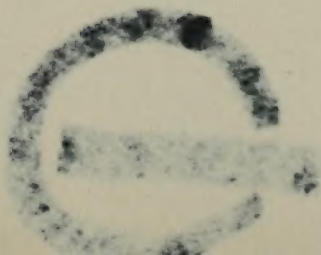
PREFACE

Although the research reported in this paper was not commenced until 1920, the interest in schoolroom tests which prompted it had its beginning in the study of the influence of varied instructions on the results of class work, a research problem which was carried on with the help and inspiration of Dr. William H. Burnham of Clark University, Worcester, Massachusetts.
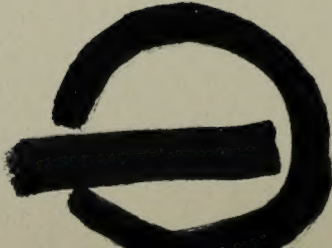
In the course of experimenting with a variety of tests the fact became clear that reading ability was highly essential to success in academic subjects. Therefore, attention was turned to tests that would indicate with reasonable accuracy the relative proficiency in reading of the individuals in any given group. Since in reading ability vocabulary must of necessity be an important factor, just how useful and convenient a group measure could be developed from vocabulary lists was the problem set for investigation.

Of the many teachers, pupils, and friends who have made this research possible only a few can be mentioned here. First of all grateful acknowledgment is made to Miss Mira H. Hall, Principal of Miss Hall's School, Pittsfield, Massachusetts, who has contributed most enthusiastic support and helpful suggestions to this work throughout all the years of study. To Dr. John F. Gannon, Superintendent of Schools, the writer is indebted for access to the Pittsfield schools and their records. Much valuable assistance was rendered by the principals and teachers who administered the first tests in the public schools and especially by the three who arranged for the recent retesting: Mr. Roy M. Strout, Mr. John A. Ford, and Mr. E. G. Bridgham. Through the courtesy of Houghton, Mifflin Company the use of the Terman list in its changed form was permitted. To her sister the writer is deeply grateful for the many hours spent in assembling and recording much of the data included in this report.

A. L. W.

April 30, 1927

CONTENTS

A Vocabulary Information Test

CHAPTER I

INTRODUCTION

The studies included in the present report have extended
over several years. Testing has been carried on in both pub-
lic and private schools in the attempt to devise a measuring
scale that is accurate enough to serve as a guide in school class-
ification, yet can be quickly administered and interpreted.

In 1920, when this vocabulary study was commenced, there
were few psychometric tests available for use in secondary
schools. Though the number of such instruments has in-
creased rapidly of late, nevertheless it has seemed worth while
to continue this work since material of this character, based
as it is upon pupils' answers, may contribute something help-
ful to a better understanding of the school's task.

In the early stages of this study material was drawn from
a variety of sources. First cancellation of nouns was tried,
using indentical paragraphs from the Boston Evening Trans-
cript. Later on the "hard directions" and "cancellation" tests
devised by Woodworth and Wells (51) were employed as time-
limit group tests. The results obtained from this experimenta-
tion indicated that reading examinations and vocabulary tests
correlated more closely with the academic marks than other
kinds of tests. This seemed to be in substantial accord with
the results of Terman's measurements of intelligence reported
in 1916. He stated that the vocabulary test had "a far higher
value than any other single test of the scale" (38). "Used with
children of English-speaking parents," he says (page 230)
"it probably has a higher value than any three other tests in
the scale." This writer's endorsement of vocabulary tests as
a mental measure was enthusiastic enough to justify further
investigation in this field. Therefore, a study of vocabulary
was commenced, using as a starting point Terman's list as re-
ported in the *Journal of Educational Psychology*, December,
1921, "Terman Vocabulary as a Group Test" (49).

One source of error in results gained by the use of the Ter-
man vocabulary was the publicity which that list had received.
For example, "homunculus, limpet, achromatic" had been

learned by pupils who saw the word lists in the *Literary Digest.* This vocabulary test, however, proved very helpful in spite of some undue familiarity with the difficult words.

The next step of the investigation was to arrange from original lists of words, selected in a similar manner, group tests of graded difficulty. Stated explicitly, it has been the purpose of this research: (1) to construct several equivalent vocabulary tests suited to the ability of grades seven through twelve; (2) to make the corresponding units of the different lists equivalent so that they may be interchangeable; (3) to provide objective scoring for the sake of speed and accuracy, and (4) to establish age and grade norms.

CHAPTER II

DEVELOPMENT OF VOCABULARY MEASUREMENT

"When you increase a student's vocabulary you increase his standing in his class." This quotation from Doran's article entitled "A Study of Vocabularies" (8) which appeared in 1907 indicates the direction in which the thought of many investigators has turned since Mueller's work in 1876. If any such correlation exists between size of vocabulary and success in school work, here may be found a new tool with which to investigate intelligence.

Much work reported during the exploratory period was unscientific. Methods were varied and records were incomplete, but out of such playing with word knowledge the more formal testing developed. In the article mentioned above Doran (8) has presented all the statistics that were then available. Other reports which furnish excellent bibliographies of vocabulary testing are those of Whipple (50), Gerlach (13), Thorndike (44), Schwesinger (34), and Kennon (21). Since, however, the purpose of this research is to standardize a word list for use in testing, no further discussion of the early efforts in the field is presented.

The second period of vocabulary investigation in this country began in 1907 when Kirkpatrick (22) reported his method of making a scale to measure the "size of vocabulary,"—what Whipple (50) terms a "vocabulary index." Of this work Whipple says (part II, page 314): "Kirkpatrick found a tendency toward positive correlation between class standing (teacher's grades) and vocabulary index." In his own classes Whipple found a correlation of +.45 ±.06 between the vocabulary index of fifty-eight college students and their grades in educational psychology.

Interest in the work of Binet and Simon which was reported in Paris in 1908 stimulated linguistic studies here. For example Terman and Childs (40) who published "A Tentative Revision and Extension of the Binet-Simon Measuring Scale of Intelligence" proposed a new list of test words based on a dictionary containing only 18,000 words, instead of an abridged dictionary of 28,000 words such as Kirkpatrick used. This list of one hundred stimulus words which Terman has discussed in "The Measurement of Intelligence" (38) is a random sampling

of the smaller dictionary. Its selection was due to the observation that the more complete the list from which the selection was made, the more it tended to reveal accidents of training rather than real intelligence. The word list of Terman and Childs was more than a "vocabulary index," for with it the authors were attempting to measure not vocabulary alone but mental age as well. Thus we see a third period in vocabulary study arriving.

In 1916 Terman, in "The Measurement of Intelligence" (38), wrote "that the value of the vocabulary test was higher than that of any other test in the mental age scale." In 1918 the same investigator (39) stated: "Vocabulary growth is constant and regular. The curve of medians of successive ages is almost a straight line." If the mental age be based on vocabulary alone, the probable error would not be greater than 9.6 months. Foreign children were behind in such a test for the first three years, but the difference disappeared by the time the mental age was twelve years.

The present attitude toward vocabulary testing seems to be the converse of Doran's suggestion. Instead of increasing the vocabulary to improve the class standing, we measure the vocabulary in the hope of discovering what the class standing ought to be.

Many forms of vocabulary tests have developed since their prognostic value was established. No battery of general intelligence examinations lacks its synonym, antonym, or sentence-completion section. Such forms as the "Association Tests" of Woodworth and Wells (51) were quickly adapted to the psychometric work and new tests were devised for the same purpose. During the last decade much attention has been directed toward developing scales that shall be at once precise and convenient to apply, thereby increasing the efficiency of this sort of mental measuring.

Gerlach (13) in 1917 published a multiple choice test of one hundred words standardized in Colorado College. This form of test is far more convenient for the examiner and, if Cannon's (5) results with four tests are typical, this "selective form," as he names it, "is equal to and possibly superior to the inventive form for use with children under fourteen years of age."

Holley (18) has published a "Sentence Vocabulary" test with forms suited in difficulty to the different school grades. There are twenty stimulus words in each list.

Inglis (19) offers a test series similar in structure, each list of which is intended "to measure the intelligent general reader's vocabulary." If, as the author suggested, subjects scored their own test blanks, less reliance is to be placed upon his figures.

Two recent tests are those of Schwesinger (34) and Kennon (21). The former, reported in "The Social-Ethical Significance of Vocabulary," samples the moral content of the child mind; the latter presents "Tests of Literary Vocabulary for Teachers of English," in which the words are "related rather specifically to the field of English literature." Both investigators, though confining their words to a specific field, bear testimony to the positive correlation between these tests and tests of general intelligence.

Enterprising educators have been eager to apply the results of the vocabulary investigation, both to the measurement of achievement in school subjects and to the prediction of future school standing. As the titles in the bibliography indicate, Thorndike has been interested in the pedagogical application of test results. In 1919 a publication of his relating to tests of intelligence (41) made reference to a "rough correspondence between scores in tests of general intelligence" and stated that they meet "the important prognostic test of predicting the limits of achievement under life conditions."

Between 1921 and 1924 this same investigator reported the results of researches in the field of vocabulary in which the aim was to discover what words should be taught to the children of each school grade (43, 44 and 45). In collaboration with Symonds (46) Thorndike has reported on the "Difficulty, Reliability, and Grade Achievements in a Test of English Vocabulary." From this study he concludes that one such test is reliable for grouping, but that four tests would be needed to rate an individual.

After all his researches in the linguistic field, it is interesting to note how much importance this author attaches to comprehension of language. In his recent book entitled "The Measurement of Intelligence" (42) the four series of tasks which constituted the symbol "CAVD" all concern responses to ideas and symbols, especially words and numbers (page 64). The list given is as follows (page 65):

"C. To supply words so as to make a statement true and sensible.

A. To solve arithmetic problems.

V. To understand single words.

D. To understand connected discourse as in oral direction or paragraph reading.

The arrangement of scoring is such as to attach equal weight to each of these four varieties of task." In chapter fourteen the meaning is discussed of scores obtained in standard intelligence examinations, which are used "for the purpose of ranking individuals according to that ill-defined trait known as intelligence." In his conclusion the author says that no one short single test can be devised which depends for its success upon all of intellect and nothing but intellect.

As to the material which is incorporated in the Standard intelligence examinations, Thorndike's criticism is thus summarized (page 404) : "Except for the speed element, then, one of these stock intelligence examinations may be regarded as a series of composites unequal in the number of their elements and undefined as to the distance between levels." This is quite suggestive of Root's (33) opinion expressed in 1922 to the effect that batteries of group tests show no careful weighting of test elements (page 291).

Each excursion into the psychometric field proves to the investigators the need of better methods of handling results. Several references included in this report indicate serious efforts which are being put forth to improve the tests themselves and to render the test results intelligible and useful to those educators and employers who possess little or no psychological knowledge. Important contributions to this work of adapting mental tests and statistical methods to pedagogical uses have been made by Haggerty (16, 52) and Otis (24-29). "The Statistical Method in Educational Measurement" (28) is a most suggestive book for teachers and school officials, presenting in clear concise form the explanation of statistical terms and methods which are essential to the interpretation of test results. While doing much to provide psychometric tools of greater precision, Otis is making a contribution which is equally important when he leads educators to make the best possible use of the tests which are now available.

Thus we find in the history of vocabulary testing evidence that word knowledge has served as a reliable measure not only of the size of vocabulary but also of the general intelligence. The most recent studies in the field are demonstrating the predictive value of such tests for vocational and educational guidance.

CHAPTER III

PRELIMINARY TESTS

In preparing material for a new test series, words were chosen from Webster's Collegiate Dictionary, third edition, 1919, which contains 97,000 words and phrases on 1,100 pages. In sampling this book one word was selected from every tenth page, thus making each list of one hundred words a random sampling of the book. Obsolete words, compounds that retain the literal meaning of their components, and technical terms were omitted.

These lists, selected from a larger dictionary than that quoted by Terman, would, according to him, reveal "accidents of previous training rather than real intelligence," but the information is equally valuable in classifying students.

Basing the test, then, on a larger vocabulary should provide lists better suited to the pupils above than to those below the seventh grade, because the older pupils of normal mentality would have profited more by previous training.

In lists of fifty these words were given on ten successive days to fifty-seven girls in Miss Hall's School, each pupil being asked to write a definition after every stimulus word. The directions were the same as those previously used for the Terman words.

The group of private school girls who defined the words in this first trial ranged in school grades from eight through twelve. The age range was from fourteen through nineteen. Half of these pupils were pursuing college preparatory courses; the rest were carrying a varied program which included more history, science, and practical arts than can be offered to college candidates.

On the basis of this trial a percentage value was assigned to each of the five hundred words, those defined correctly by all subjects being rated as 100 per cent words and words missed by all as zero words. This entire list may be found in Table I.

TABLE I
Stimulus Words

Five hundred words graded in accordance with the percentage of correct definitions which they elicited from fifty-seven pupils in Miss Hall's School and arranged in an order of increasing difficulty.

100—always, April, brainless, caller, century, cheer, cyclone, debate, forehead, palace, sober, speaker, surgeon, tomorrow, watermelon, whipped.
<div align="right">16 words</div>

98-99—adore, ankle, bear, cushion, diary, dollar, grey, happy, hockey, incomplete, luggage, new, post, regard, scheme, step, swelling, sweater tangerine, tomahawk, visit.
<div align="right">21 words</div>

96-97—alchoholic, bigger, bugle, camouflage, cause, chatterer, cretonne, enclose, friend, illumination, lawless, major, old, pen, republish, salmon, sanitarium, sequel, snow, spur, sugar, tapestry, travel, understand.
<div align="right">24 words</div>

94-95—faithless, flexible, guard, hazy, horizon, index, knotted, lead, method, month, orchard, peg, queer, reformation, rough, top.   16 words

92-93—boast, cozy, emperor, endorse, frigid, jest, kerosene, muscular, reach, scarf, season, settee, sheriff, skeleton, tight, tire, vault, ventilation, weasel, whole.
<div align="right">20 words</div>

90-91—afterwards, axle, balance, bank, cobbler, decide, dickey, dingy, footless, game, ginger, hominy, jockey, microscope, nestling, represent, rhododendron, scalp, strike, strode, struggle, terrorize, touching.
<div align="right">23 words</div>

88-89—altar, ambition, brier, casino, charmer, crook, equator, geography, ignorantly, massive, memory, menagerie, reseat, ripple, serviceable, sovereign, summons, trust.
<div align="right">18 words</div>

86-87—confirm, diet, holding, kneel, laugh, plunderer, pretender, rate, teach.
<div align="right">9 words</div>

84-85—ashes, atmosphere, freightage, menace, plotter, prairie, preserver, puncher, stalwart, switch.
<div align="right">10 words</div>

82-83—assure, beauteous, bedridden, fagot, insult, minuet, passport, plainsman, plateau, poetical, politically, purchasable, ruffle, shrewd, sulphur, writhe.
<div align="right">16 words</div>

80-81—betterment, fit, fray, intense, lick, mercy, mingler, shine, staghound.
<div align="right">9 words</div>

78-79—certain, electrical, monoplane, pavilion, photoprint, restless, utmost, vengeance.
<div align="right">8words</div>

76-77—blowing, cute, disk, nude.   4 words

74-75—arteries, fiction, historicalness, larynx, monkhood, pyramid, snappish, steadfast, tributaries, vertical.
<div align="right">10 words</div>

72-73—Gloria, inflexibility, notwithstanding, reëngage, subdivide.
<div align="right">5 words</div>

70-71—Biblist, billow, magnetism, perpendicular, prong.   5 words

68-69—bunting, counsel, bravery, fifer, heifer, hemisphere, marquis, rattly, series, shamefaced, shoer, shyly, tallish, trotter.   14 words

TABLE I—continued

66-67—bellows, conditional, damask, eradication, ferret, geyser, harem, jubilee, reliance, turbulent.   10 words

64-65—carbuncle, drudge, incapability, Olympian, palate, pewter, poser, skeptic.   8 words

62-63—departmental, grace, longitude, may, recession.   5 words

60-61—essence, excess, fibrous, jollity, limited, loot, manakin, mentality, officiate, warrantable.   10 words

58-59—apportion, connective, coupling, especial, excommunicate, guilt, hundredth, pulley, quintet.   9 words

56-57—acid, doer, dubbing, impeachment, nape, radical, unreel.
                                                                   7 words

54-55—barrage, Celt, enforce, glazier, Harpy, hussy, Irishism, latitude.
                                                                   8 words

52-53—abash, ample, fledgling, marrow, mosaic, paternity, saintship, vertebra, virginity.   9 words

50-51—abusive, ban, benumb, canine, complement, consign, disfiguration, Gallic, irreducible, linseed, martyrize, obnoxious, serum.   13 words

48-49—bluet, inferiority, isthmus, tensity.   4 words

46-47—decade, dyspeptic, forgetful, idiosyncrasy, spar.   5 words

44-45—aday, ascendable, Hellene, imbibe, logicality, principle, transient, vouchsafe.   8 words

42-43—artificer, carmine, interceptor, longevity, mimeograph, obscurity, pip, rosemary.   8 words

40-41—boycott, crinoline, Magi, Parthenon, unanimity, venom, worldliness.   7 words

38-39—dissertation, farcical, gossamer, prophecy, shimmer.   5 words

36-37—elapse, equilateral, intelligible, Madras, poniard.   5 words

34-35—acquittal, agglomeration, augment, ballardry, cartilage, contrabandist, laky.   7 words

32-33—ballast, carouse, commodity, contestable, eject, equinox, hoo, infusion, perforce, phase, proteid, salivary, syndicate.   13 words

30-31—humanist, inkiness, magisterial, presbyter, selfhood, supplement.
                                                                   6 words

28-29—albumen, chrysalis, declivity, denominational, reprieve, hyperbolic.
                                                                   6 words

26-27—cohesive, meridian, narcotic.   3 words

24-25—cloven, confabulation, esophagus, lionize, Pantheon   5 words

22-23—append, capillaries, paroxysmal, pelvis, vellum.   5 words

20-21—auspicious, biped, contrite, discreate, dominie, eely, effusion, elicit, expiation, fiord, lyceum, phlegm, postulate, prolocutor, seconder.
                                                                  15 words

TABLE I—continued

18-19—archives, Chinook, crux, kiln, pall, purloin, scavenger.   7 words

16-17—alimentary, automata, grail, jamb, quadrant, stockinet, umber, weevil.   8 words

14-15—circumfuse, copperas, defloration, epiglottis, incisors, penitential, ping, thane, thorax.   9 words

12-13—actionable, cinquefoil, foment, gaud, plantain, saline.   5 words

10-11—calyx, comfit, erratum, foh, irruption, mummer, Ostrogoth, sagittary.   8 words

8-9—aloes, bullpout, canister, Cecrops, Galantine, pancreatic, pylorus.   7 words

6-7—Agnus, coadunate, Cotswold, heronry, nuncio, oscular, seine, tonicity.   8 words

4-5—anticly, dulcimer, ethnology, flagellate, incipient, lignite, Scorpio, shote.   8 words

2-3—anomalism, claque, clavicle, deprecate, escarpment, fennel, flexors, quiesce, secularize.   9 words

0-1—abigail, cardoon, cicely, commensal, despumate, exegesis, facet, felly, gibbous, hadji, kakapo, Kremlin, ligate, mist, mistral, paddy, petaloid, sistrum, spheroid, springe, tricennial.   21 words

With the exception of facet, gibbous, mist, and spheroid, all words in the last paragraph rated zero for the school where the scale was made.

The distribution of these words when they were arranged in six per cent intervals is shown by Table II. In this table the most difficult words are included between zero and five per cent.

TABLE II

Frequency Distribution of 500 Stimulus Words Classified by Per Cent of Correct Definitions Elicited from Fifty-seven Girls

Per cent Right	No. of Words	Per cent Right	No. of Words
96–100	61	42–47	21
90–95	59	36–41	17
84–89	37	30–35	26
78–83	33	24–29	14
72–77	19	18–23	27
66–71	29	12–17	23
60–65	23	6–11	23
54–59	24	0–5	38
48–53	26		

Inspection of these two tables reveals the fact that, of the five hundred words, one hundred and ninety-four rated over seventy-five per cent, while one hundred eleven rated less than twenty-four per cent. The upper quarter from seventy-five to one hundred per cent contained as many words as the two middle quarters added together. When grouped in steps of two per cent, there were twenty-one words in the zero to one interval and sixteen words that were defined correctly by all the girls in the group. The words were unevenly grouped along the scale, but there were enough at every interval to furnish three equivalent fifty-word lists with a difference in difficulty of two per cent between successive test units.

Accordingly multiple choice tests were constructed from these definition lists. The arrangement of stimulus words from easy to hard accords with the percentage rating derived from the first tryout and the confusions presented are some of the errors which appeared on those papers. For the definitions single words have been used in preference to phrases in order to equalize the amount of reading required for the different lines. Owing to the small number of words in one interval only three word lists were made. Had larger intervals been employed, a better selection of words could probably have been made with no more loss in accuracy of grading than the change to the multiple choice form of test occasioned.

It is interesting to note that the foregoing tables indicate also what proportion of the 97,000 words in the Collegiate Dictionary were known to girls of the type described above. Since each list of one hundred words was a random sampling of the entire book, a frequency of sixty-one in Table II indicates that approximately twelve per cent of the words in the list or a total of 11,640 words was understood by ninety-six per cent of these girls, while only seven per cent or 6,790 were wholly unknown to this group.

CHAPTER IV

THE MULTIPLE CHOICE TESTS OF 1922

When the one hundred fifty stimulus words which stood at regular intervals on the scale had been arranged in three lists of regularly increasing difficulty, three confusions for each word were selected from the definitions obtained in the first tryout. The correct definition was taken from the synonyms given in Webster's Collegiate Dictionary. Thus each line of the multiple choice test consisted of five words: one stimulus, one exact definition, and three confusions.

In 1922 when a mental survey was carried on in the public schools of Pittsfield, Massachusetts, these three multiple choice tests were tried in grades seven to twelve. By vocabularies, the number of records secured was as follows:

Vocabulary   I   386 tests
Vocabulary II   765 tests
Vocabulary III 1351 tests

Total   2502 tests

Each vocabulary was printed on a single page with space in the heading for the personal data and directions. The latter read:

DIRECTIONS. Look at the line of words across the page reading sparrow—fish, water, bird, animal. Think what the first word means, draw a line under that one of the four words after it which best gives the meaning of the first word. Do the same way with the second line of words, and so on with all the succeeding lines until you are told to stop.

Examples: sparrow— fish, water, <u>bird,</u> animal
toil— play, <u>labor,</u> stop, prosper

The three vocabularies are reproduced in Table III which follows:

TABLE III
Vocabularies I, II, III

	I				
1	always	now	before	hereafter	ever
2	tomahawk	battle-ax	horn	animal	bird
3	camouflage	upset	hide	disguise	steal
4	peg	knob	support	bolt	hook
5	muscular	strong	energy	strength	tall
6	cobbler	stone-worker	shoe-mender	salesman	merchant
7	ambition	height	future	eagerness	accomplishment
8	confirm	conform	imprison	confine	establish
9	prairie	desert	wilderness	meadow	forest
10	fagot	cigarette	bind	cipher	herb
11	fray	fight	party	edge	foe
12	monoplane	one-cylinder	plane	wing	airship
13	cute	bow-legged	adorable	shrewd	sweet
14	fiction	rubbing	invention	pressure	sensation
15	notwithstanding	however	because	besides	unless
16	billow	wave	roar	noise	tree
17	heifer	dash	sheep	cow	bull
18	geyser	goose-herd	spring	volcano	garment
19	drudge	weary	grudge	burden	toil
20	longitude	width	latitude	line	length
21	loot	jeer	plunder	shout	lute
22	pulley	rope	wire	tool	cable
23	acid	sour	bitter	slip	strong
24	barrage	wall	gun-fire	rain	men
25	abash	confuse	bashful	shy	scared
26	serum	syrup	rinse	wash	fluid
27	bluet	color	bird	innocence	grass
28	decade	old	period	rotten	century
29	principle	chief	head	money	source
30	obscurity	abstruseness	hidden	safe	crooked
31	boycott	steal	catch	avoid	house
32	dissertation	rebuke	disquisition	desert	distraction
33	elapse	pass	part	space	faint
34	acquittal	pardon	leaving	punishment	discharge
35	perforce	perhaps	perchance	necessarily	unless
36	supplement	request	supple	stubstitute	appendix
37	reprieve	scold	delay	reprimand	repay
38	narcotic	poison	sleepy	stimulant	stimulating
39	pantheon	animal	colonnade	building	amphitheatre
40	append	bear on	suspend	attach	threaten
41	postulate	condition	object	rule	articulate
42	archives	onions	annals	vegetables	herb
43	grail	altar	bird	box	chalice
44	incisors	instruments	scissors	cuts	teeth
45	plantain	field	farm	weed	vegetable
46	comfit	comforter	puff	ease	preserve
47	canister	candle	box	singer	lawyer
48	seine	net	river	name	wool
49	lignite	muscle	coal	heat	light
50	claque	sound	talk	applauders	fake

TABLE III—continued

II

#				
1	cushion	chair	pillow	shoe	support
2	friend	liking	person	love	intimate
3	index	list	book	card	box
4	tire	tare	wheel	exhaust	wheat
5	dingy	small	soiled	bare	stingy
6	massive	weighty	crowd	plenty	full
7	plunderer	blunderer	diver	pillager	plunger
8	freightage	car	train	send	cargo
9	ruffle	untidy	dress	lake	wrinkle
10	shine	bright	brilliant	sheen	reflect
11	pavilion	covering	walk	plane	balcony
12	nude	rude	naked	prig	fop
13	vertical	horizontal	upright	straight	oblique
14	inflexibility	pliable	immovable	bending	stiffness
15	perpendicular	hanging	even	vertical	parallel
16	shamefaced	bashful	guilty	disgraceful	boy
17	bellows	roars	moving	wind	fire
18	palate	neck	throat	bed	roof
19	grace	dainty	slim	charm	fairy
20	jollity	person	gayety	joyful	funny
21	coupling	separating	tin	joining	verse
22	dubbing	naming	bluffing	fooling	mixing
23	latitude	length	breadth	height	longitude
24	marrow	small	depths	thin	essence
25	linseed	varnish	cotton	oil	flax-seed
26	tensity	strong	thickness	strain	deep
27	forgetful	person	heedless	forget	memory
28	vouchsafe	grant	bet	ascertain	exclaim
29	rosemary	beads	perfume	school	name
30	Magi	dwarf	spirit	sorcerer	bird
31	prophecy	think	fulfill	foresee	prediction
32	intelligible	knowing	wise	bright	comprehensible
33	augment	discussion	help	increase	debate
34	infusion	tincture	excite	thrust	insert
35	inkiness	inky	blackness	black	smudged
36	chrysalis	pupa	stone	glass	flower
37	meridian	person	latitude	culmination	sea
38	cloven	condensed	woven	stuck	split
39	capillaries	blood-vessels	nourishment	worms	intestines
40	discreate	careful	annihilate	wise	sly
41	purloin	buy	prolong	meat	steal
42	alimentary	nutritious	ailing	beginning	primary
43	thane	two	tribe	chief	wood
44	foment	waste	intrigue	spoil	decay
45	irruption	outbreak	overflow	inroad	rise
46	pancreatic	nerve	sweetbread	universal	panicstricken
47	nuncio	messenger	speak	pronounce	announcement
48	dulcimer	sweet	soft	gentle	instrument
49	clavicle	hollow	throat	collarbone	shoulder
50	kakapo	parrot	nut	tribe	medicine

TABLE III—continued

III

1	scheme	plan	construct	quarrel	teach
2	signal	time	place	token	way
3	guard	house	protect	fear	coat
4	emperor	position	power	priest	ruler
5	footless	broken	clever	inapt	whole
6	trust	gift	faith	truth	value
7	teach	instruct	learn	school	acquire
8	plotter	conspirator	plan	opposition	blotter
9	beauteous	beauty	plentiful	quality	handsome
10	intense	ceaseless	lot	earnest	entire
11	certain	assure	surely	prove	reliable
12	disk	cylinder	plate	corner	sphere
13	tributaries	pay	gathering	subjects	gifts
14	subdivide	divide again	half divide	part	divide under
15	magnetism	attractive	drawn	enlarge	power
16	counsel	meeting	advice	group	law
17	ferret	nail	bird	animal	boat
18	harem	wives	palace	man	custom
19	recession	reform	meeting	beginning	withdrawal
20	excess	outlet	overplus	exit	access
21	especial	sure	particular	very	reserved
22	radical	political	extreme	exactly	bright
23	enforce	compel	strengthen	aid	increase
24	ample	plenty	lot	wide	creature
25	benumb	unfeeling	unconscious	freeze	deaden
26	inferiority	interference	importance	lowliness	superior
27	idiosyncrasy	peculiarity	idiotic	folly	jest
28	imbibe	bribe	force	induce	imbue
29	artificer	lawyer	workman	musician	scholar
30	venom	snake	zeal	reptile	poison
31	farcical	falsely	joke	unreal	facial
32	Madras	city	cloth	islands	grapes
33	cartilage	load	gristle	bone	bullet
34	syndicate	signify	tell	indicate	council
35	presbyter	desk	cushion	elder	church
36	declivity	decline	chasm	condition	abyss
37	cohesive	plaster	unity	following	united
38	esophagus	vegetable	gullet	wind-pipe	hill
39	vellum	volume	wish	parchment	wall
40	auspicious	fortunate	odd	suspicious	apparent
41	pall	mantle	dish	coffin	body
42	weevil	rat	beetle	cootie	bird
43	defloration	decoration	extraction	desecration	blossoming
44	gaud	loud	glad	urge	trinket
45	mummer	buffoon	silencer	joker	speak
46	aloes	metals	plant	polish	iron
47	coadunate	coax	help	accompany	combine
48	Scorpio	Roman	spider	constellation	philosopher
49	flexors	instrument	muscles	scissors	glands
50	springe	leap	spray	jump	snare

From the results of this trial a rating was computed for each line in the tests, the rating of each being determined by the percentage of correct answers it elicited. Thereafter each line of the test was treated as a unit; no change whatever was made either in the definitions or the confusions, since the relative difficulty of a line depends as much on the responses offered as on the stimulus word.

The values assigned the lines in the three lists of this first multiple choice series are summarized in Table IV. There it is seen that Vocabulary II proved to be distinctly more difficult than the other lists. Since this had not been true of the separate words when they were first rated in Miss Hall's School, it could not be determined without further investigation, whether this difference was due to the stimuli, the responses or the group tested.

TABLE IV

Showing the Rating of One Hundred Fifty Test Units When Presented to Public School Pupils in Multiple Choice Tests

Per Cent Right	Vocabulary I	Vocabulary II	Vocabulary III
100	0	0	0
90–99	2	1	4
80–89	2	1	3
70–79	3	4	2
60–69	5	2	3
50–59	6	5	8
40–49	4	6	10
30–39	11	9	6
20–29	9	9	8
10–19	5	8	5
0–9	3	5	1

A comparison of the three vocabularies was made to determine the proportion of easy, average, and difficult lines included in each list. The ratings were interpreted as indicated in Table V.

TABLE V

Showing the Number of Easy, Average, and Difficult Lines in the Three Multiple Choice Tests

Rating Per Cent	Easy 70-100	Average 30-69	Difficult 0-29	Total
Vocabulary I	7	26	17	50
Vocabulary II	6	22	22	50
Vocabulary III	9	27	14	50

The first and third lists are nearly equivalent but 44 per cent of the test units in the second list proved to be difficult for grades 7, 8, and 9. Several familiar stimulus words were missed when followed by more than one response that approximated the desired definition, pupils failing to select the best meaning. This fact was of little consequence, however, in this research since the discriminating value of lines rather than of separate words was sought. The individual words were studied very carefully when the same mistake was made by a large number of subjects.

After the publication of "The Teacher's Word Book" by Thorndike (43) the 1300 words of these tests were compared with the experimental lists in order to discover the relative difficulty of the responses as well as of the stimulus words. Many words shown by the latter to be difficult are so rare as not to be included in Thorndike's Word-Book. It was surprising, however, that some of the easy words did not appear in this list, but these lacks may be explained by the fact that Thorndike's work was based to some extent upon concordances. Thus "swaddle," a word included in Terman's vocabulary is omitted by Thorndike, although it is understood by many children below the eighth grade, because of its occurrence in a familiar Bible story.

A tabulation of the displacement of 100 words shows the changes in rank order which occurred when these words were presented to public school children in a multiple choice test. The number of such changes is recorded in Table VI. Here the first column represents the difference between the rank or-

TABLE VI

Showing the Displacement of Words in the Scale When Rated According to Test Records of the Multiple Choice Tests

Rank Order Change Not over	Voc. II (765 Cases) Word-Freq.	Per Cent	Voc. III (1351 Cases) Word-Freq.	Per Cent
4	20	40	24	48
9	32	64	33	66
14	38	76	44	88
19	41	82	47	94
24	47	94	48	96
29	47	94	49	98
34	48	96	50	100
39	50	100		
Average displace-placement	12.7 points		10.5 points	

der of the stimulus word in the original word list and that of the corresponding line in the multiple choice test.

More change in the rating of the test units was manifest in Vocabulary II than in Vocabulary III. This accords with the fact that this list contained more difficult words than the other and would therefore be more subject to the "accidents of training" mentioned by Terman.

No time limit was placed on these tests, but each examiner reported the time required by her grade to complete a list. According to their reports ten minutes was a long enough period to allow for one vocabulary. In that time pupils could show very fairly how much of the task was suited to their ability. If many words were omitted or answered incorrectly, this indicated that the test was too difficult for the child, because no one was compelled by lack of time to leave the task unfinished.

Several factors operated to disturb the grading of the test items. First: The subjects who furnished the first definitions were fifty-seven girls ranging in school grades from eight to twelve. The average intelligence rating of this group was superior. They were accustomed to test conditions. The public school group was more heterogeneous in both chronological age, eleven to twenty-two years, and school grade, seven to post-graduate. Each principal had been a law unto himself in the matter of promotion and classification with the result that all grades of intelligence from border-line to superior might be found in any one class. No psychometric tests had been previously used in the schools.

Second: The task in its nature is different. To give a definition requires acquaintance with only two words, while selecting the best definition out of four possible ones calls for a larger vocabulary. This apparent increase in difficulty is, however, partially balanced by the suggestion offered in the definitions presented. It seems easier to underline a word that is given than to think of a suitable word to write.

Third: Errors are frequently made in the multiple choice which would not occur if free definitions were given. For "fiction," "rubbing" was often chosen; for "rosemary," "rosary"; for "springe," "leap" or "jump." These errors are due to careless reading.

Sometimes a definition is chosen which sounds like the stimulus word, regardless of the part of speech or meaning. The

s on "aloes" seemed to lead to the marking of "metals" which has a similar plural form. "Mummer" was often defined as "silencer," though "buffoon" was the first definition given.

In some cases a tendency appeared to mark the first word that was at all satisfactory rather than to find the best definition. The last line in Vocabulary III illustrates this sort of error. Here "leap," which was the first meaning offered, was marked more frequently than "jump," which was the third response word, though both words were equally incorrect.

A tabulation of one thousand eight errors made by the pupils of Miss Hall's School on the last half of page two showed that among the columns errors were distributed as follows:

TABLE VII

Showing the Frequency Distribution of One Thousand Eight Errors among the Four Columns of the Test

Column	1	2	3	4
Frequency of errors	303	283	250	172

Only 17.2 per cent of these errors appeared in column four, while 30.3 per cent of them occurred in the first column. Therefore, the behavior of subjects who are guessing is not controlled wholly by chance. Other factors enter into the situation. Some pupils stated that they marked columns in rotation according to a system of their own because they knew the answers were distributed.

The right answer received no undue attention when in the first column. On the same half page two columns, the first and the third, had five correct answers each, the column one definitions occurring between lines twenty-seven and forty-nine, those of column three between thirty-one and forty-seven. In the former the slightly greater range of difficulty could hardly account for the fact that only eighty-seven right answers were found in the first column, while one hundred eighteen were correct in column three. There is a tendency to overlook the correct answer when it is in the first column. From the five hundred sixty-six correct answers found on this half page, the per cents found in the respective columns are as follows: First, sixteen; second, thirty-three; third, twenty-one; fourth,

thirty. The correct answers are marked most frequently in the second column and the errors in the first.

The distribution of percentage values of lines showed that the test units needed thorough rearranging. Examination of the errors on the test blanks suggested changes which might lead to more accurate results.

To eliminate errors due to careless reading it seemed advisable to have all answers underlined and also to have their numbers written in parentheses provided at the ends of the lines, thus requiring all subjects to read the whole line. This doubling of the response might add to the time required, but that was of little consequence if it increased the accuracy of reading.

Before excluding the most difficult words from the lists it seemed best to try them again in the senior high school using a more convenient test blank, more explicit directions, and clearer type. The Berkshire Word Test was constructed for this purpose.

TABLE VIII

The

BERKSHIRE WORD TEST

Form A.   For Grades 7-12

Arranged and standardized in the schools of
Pittsfield, Massachusetts

⚬⚬⚬

Name ....

Age last birthday........*Date of test*......................

Date of birth................*School*.....................

City .... *Grade*.............

⚬⚬⚬

This is a test to see how many words you know. You will find on the following pages columns of words like the sample words given below. You see that the words in the first column are printed in heavy type. After each of these words **four meanings** are given in light face type in the columns numbered 1, 2, 3, 4.

Read carefully the first word in heavy type then look carefully at the four meanings in light face type **on the same line.** Underline the **one** word which best gives the meaning of the word in the heavy type. Then in the parentheses at the right place number of the column where you underlined the meaning.

For example see the sample words below:

	1	2	3	4	
sparrow	fish	water	<u>bird</u>	animal	(3)
toil	play	<u>labor</u>	stop	prosper	(2)

Do the same for all the words in heavy type. When you have finished page 2, turn to page 3 and so on until you have marked the meaning of every word in heavy type.

PAGE 2

		1	2	3	4	
1	scheme	plan	construct	quarrel	teach	( )
2	guard	house	protect	fear	coat	( )
3	emperor	position	power	priest	ruler	( )
4	ample	plenty	lot	wide	creature	( )
5	trust	gift	faith	truth	value	( )
6	teach	instruct	learn	school	acquire	( )
7	plotter	conspirator	plan	opposition	blotter	( )
8	especial	sure	particular	very	reserved	( )
9	enforce	compel	strengthen	aid	increase	( )
10	beauteous	beauty	plentiful	quality	handsome	( )
11	inferiority	interference	importance	lowliness	superior	( )
12	excess	outlet	overplus	exit	access	( )
13	Madras	city	cloth	islands	grapes	( )
14	ferret	nail	bird	animal	boat	( )
15	radical	political	extreme	exactly	bright	( )
16	footless	broken	clever	inapt	whole	( )
17	intense	ceaseless	lot	earnest	entire	( )
18	recession	reform	meeting	beginning	withdrawal	( )
19	subdivide	divide again	half divide	part	divide under	( )
20	venom	snake	zeal	reptile	poison	( )
21	harem	wives	palace	man	custom	( )
22	tributaries	pay	gathering	subjects	gifts	( )
23	presbyter	desk	cushion	elder	church	( )
24	disk	cylinder	plate	corner	sphere	( )
25	syndicate	signify	tell	indicate	council	( )
26	counsel	meeting	advice	group	law	( )
27	declivity	decline	chasm	condition	abyss	( )
28	artificer	lawyer	workman	musician	scholar	( )
29	benumb	unfeeling	unconscious	freeze	deaden	( )
30	weevil	rat	beetle	cootie	bird	( )
31	vellum	volume	wish	parchment	wall	( )
32	farcical	falsely	joke	unreal	facial	( )
33	magnetism	attractive	drawn	enlarge	power	( )
34	cartilage	load	gristle	bone	bullet	( )
35	signal	time	place	token	way	( )
36	certain	assure	surely	prove	reliable	( )
37	idiosyncrasy	peculiarity	idiotic	folly	jest	( )
38	flexors	instrument	muscles	scissors	glands	( )
39	cohesive	plaster	unity	following	united	( )
40	coadunate	coax	help	accompany	combine	( )
41	imbibe	bribe	force	induce	imbue	( )
42	auspicious	fortunate	odd	suspicious	apparent	( )
43	pall	mantle	dish	coffin	body	( )
44	aloes	metals	plant	polish	iron	( )
45	aesophagus	vegetable	gullet	windpipe	hill	( )
46	Scorpio	Roman	spider	constellation	philosopher	( )
47	defloration	decoration	extraction	desecration	blossoming	( )
48	gaud	loud	glad	urge	trinket	( )
49	mummer	buffoon	silencer	joke	speak	( )
50	springe	leap	spray	jump	snare	( )

PAGE 3

	1	2	3	4	
1 orange	tropical	yellow	round	fruit	( )
2 bonfire	fire in a field	sticks	leaves	fire	( )
3 straw	hay	mattress	stalk	plant	( )
4 roar	loud	lion's noise	growl	yell	( )
5 haste	speed	quickly	fast	hurried	( )
6 afloat	stays	drifting	on top	boat	( )
7 guitar	has 4 strings	has 5 strings	6 strings	8 strings	( )
8 mellow	fruit	sound	yellow	ripe	( )
9 impolite	disagreeable	cross	rude	rough	( )
10 plumbing	fixture	piping	wash-stand	water-system	( )
11 noticeable	conspicuous	seeing	attractive	unusual	( )
12 muzzle	cover	for dog	cap	covering	( )
13 quake	scared	move	shake	frighten	( )
14 reception	tea	entertainment	ball	party	( )
15 majesty	dignity	royal	impressive	great	( )
16 treasury	place	bank	deposit	money	( )
17 misuse	neglected	treatment	abuse	ill-treated	( )
18 crunch	crumple	grind	grating	avoid	( )
19 forfeit	imitation	lost	paid	fine	( )
20 sportive	sporty	sport	athletic	playful	( )
21 apish	large	silly	elfish	clownish	( )
22 snip	brat	bite	bit	snob	( )
23 shrewd	clever	coarse	frank	angry	( )
24 repose	down	rest	spot	sleep	( )
25 peculiarity	singularity	odd	different	original	( )
26 conscientious	independent	energetic	capable	faithful	( )
27 charter	rent	chart	document	maps	( )
28 coinage	change	money	treasury	introduction	( )
29 dilapidated	old	tired	torn	ruined	( )
30 promontory	projection	promise	promote	path	( )
31 avarice	evil	hatred	greed	proud	( )
32 gelatinous	stiff	viscous	sticky	jelly	( )
33 drabble	draggle	drip	dabble	dribble	( )
34 philanthropy	money	saving	industry	charity	( )
35 irony	grey	harsh	sarcasm	cold	( )
36 embody	vision	make	pile	organize	( )
37 swaddle	wrapped	swathe	dress	sacred	( )
38 exaltation	elevation	joy	pleasure	rejoicing	( )
39 infuse	mix	stir	inspire	sort	( )
40 selectman	churchman	jury	town-officer	best-man	( )
41 declivity	slope	cliff	steep	bluff	( )
42 laity	ministers	layman	purity	officers	( )
43 fen	park	grove	scenery	swamp	( )
44 sapient	absorbing	hoping	satisfied	wise	( )
45 cameo	pin	wax	gem	coral	( )
46 theosophy	science	philosophy	etymology	doctrine	( )
47 precipitancy	headlong	speed	retarding	continue	( )
48 paleology	archeology	zoology	ethnology	geology	( )
49 homunculus	gnome	dwarf	man	troll	( )
50 limpet	bird	slow	shell-fish	sluggish	( )

PAGE 4

		1	2	3	4	
1	cushion	chair	pillow	show	support	( )
2	index	list	book	card	box	( )
3	coupling	separating	tin	joining	verse	( )
4	capillaries	blood-vessels	nourishment	worms	intestines	( )
5	freightage	car	train	send	cargo	( )
6	friend	liking	person	love	intimate	( )
7	bellows	roars	moving	wind	fire	( )
8	plunderer	blunderer	diver	pillager	plunger	( )
9	massive	weighty	crowd	plenty	full	( )
10	ruffle	untidy	dress	lake	wrinkle	( )
11	vertical	horizontal	upright	straight	oblique	( )
12	jollity	person	gayety	joyful	funny	( )
13	forgetful	person	heedless	forget	memory	( )
14	inkiness	inky	blackness	black	smudged	( )
15	perpendicular	hanging	even	vertical	parallel	( )
16	prophecy	think	fulfill	foresee	prediction	( )
17	latitude	length	breadth	height	longitude	( )
18	grace	dainty	slim	charm	fairy	( )
19	pavilion	covering	walk	plane	balcony	( )
20	tensity	strong	thickness	strain	deep	( )
21	dubbing	naming	bluffing	fooling	mixing	( )
22	inflexibility	pliable	immovable	bending	stiffness	( )
23	nuncio	messenger	speak	pronounce	announcement	( )
24	rosemary	beads	perfume	school	name	( )
25	vouchsafe	grant	bet	ascertain	exclaim	( )
26	nude	rude	naked	prig	fop	( )
27	clavicle	hollow	throat	collarbone	shoulder	( )
28	dingy	small	soiled	bare	stingy	( )
29	shamefaced	bashful	guilty	disgraceful	boy	( )
30	discreate	careful	annihilate	wise	sly	( )
31	cloven	condensed	woven	stuck	split	( )
32	purloin	buy	prolong	meat	steal	( )
33	intelligible	knowing	wise	bright	comprehensible	( )
34	Magi	dwarf	spirit	sorcerer	bird	( )
35	chrysalis	pupa	stone	glass	flower	( )
36	thane	two	tribe	chief	wood	( )
37	augment	discussion	help	increase	debate	( )
38	tire	tare	wheel	exhaust	wheat	( )
39	alimentary	nutritious	ailing	beginning	primary	( )
40	dulcimer	sweet	soft	gentle	instrument	( )
41	kakapo	parrot	nut	tribe	medicine	( )
42	palate	neck	throat	bed	roof	( )
43	infusion	tincture	excite	thrust	insert	( )
44	linseed	varnish	cotton	oil	flaxseed	( )
45	pancreatic	nerve	sweetbread	universal	panicstriken	( )
46	marrow	small	depths	thin	essence	( )
47	foment	waste	intrigue	spoil	decay	( )
48	shine	bright	brilliant	sheen	reflect	( )
49	meridian	person	latitude	culmination	sea	( )
50	irruption	outbreak	overflow	inroad	rise	( )

THE END

CHAPTER V

The Berkshire Word Test

After careful review of the facts discovered in preliminary tests, The Berkshire Word Test was compiled in the autumn of 1926 (Table VIII). Page one contains personal data, directions, and two sample lines. The three subsequent pages contain two hundred fifty words each; fifty stimulus words and two hundred responses. Page two is a rearrangement of Vocabulary III of the 1922 series which had averaged 18.08 ± 4.94; page four is the old Vocabulary II which averaged 17.2 ± 5.54; page three is one of the Terman tests (38) arranged as a multiple choice test by the same method that was employed in making the lists above.

The Berkshire Word Test is only approximately reproduced here, the original sheet and print being larger, the type heavier, and italics being used for the response words.

The standard Terman vocabulary, reprinted here with the permission of Houghton Mifflin Company, publishers, was included in the test series to furnish a relative measure of difficulty and to indicate the effect of changing such a test from the free definition to the multiple choice type.

To insure more careful reading, a double response was required in this test, underlining and entering a number in the parentheses at the right, as described on page one of the test which is reproduced in Table VIII.

Group tested. "Form A" of The Berkshire Word Test was given to one hundred pupils of Miss Hall's School and one thousand three hundred forty-seven pupils of the schools of Pittsfield, Massachusetts. In the public schools the tests were administered in all classes by the respective teachers. The tests in Miss Hall's School were conducted by the investigator, who supervised all correcting and scoring.

The girls of Miss Hall's School varied in chronological age from thirteen to eighteen years, the average being about sixteen years. The average IQ was 120. In academic grouping these subjects were distributed over five grades, nine through thirteen, being about evenly divided between the college preparatory and general courses. All were accustomed to mental tests. Though the members of the group varied widely, nearly eighty per cent of them were included in the following age

groups: fifteen, sixteen, and seventeen. Only fifteen girls were over seventeen years and fourteen were less than fifteen years of age.

Though ninety-three members of the school were boarding pupils, the test results are not affected greatly by uniformity in living conditions, which are balanced by diversity in the home environment and earlier training. One was born in France; one, in Honolulu; one, in Australia. Among the others are representatives of all parts of the land; the New England States, the Middle Atlantic, mid-western, and southern.

The public school pupils represented two schools of senior high school grade, "The Pittsfield High School" and "The Commercial High School," and one, Junior High School. Since no regular program of intelligence testing was carried on in these schools, many pupils from rural schools were not test wise; but the majority of those trained in the city schools were experienced with such tests.

The accompanying table indicates the total number of cases tested in each age and grade group. In this and all succeeding tables the grade number refers to both semesters of the grade considered unless otherwise marked. The age refers to the number of years completed at the time of the test. In the group of twelve-year-olds all were included who were twelve but not thirteen; that is, those whose ages fell between twelve years and twelve years, eleven months.

In Table IX the grade-age distribution of the 1347 public school pupils is given. The large number of pupils in grade

TABLE IX
Grade-Age Distribution of 1347 Cases in Public Schools

Age Grade	7	8	9	10	11	12	Totals
20					2	4	6
19				1	4	19	24
18				11	22	70	103
17			1	32	66	97	196
16		1	22	151	98	26	298
15	7	6	31	215	36	2	297
14	21	34	60	71	1	1	188
13	44	78	12	3			137
12	60	15	1				76
11	20						20
10	2						2
Totals	154	134	127	484	229	219	1347

10 is due to the fact that 10B was tested in both semesters of the year. One half of grade 12 was retested near the close of the year, thus increasing the number of cases recorded in that grade. The age groups show great overlapping, age fourteen appearing in every grade tested. In grades 11 and 12 the ages varied from fourteen to twenty years. The age groups from thirteen to eighteen are large enough to be fairly representative of the Pittsfield school population. In all grades enough pupils were tested to furnish significant grade measures.

Methods of conducting the tests. Two methods were employed in conducting the tests. The first was like that used in the "Otis Self-Administering Tests of Mental Ability" (26). When the personal data blanks on the first page had been filled, the pupils were told to read the directions carefully. Opportunity was given them to ask questions about any part which was not clear to them, after which they turned to page two and worked without interruption for thirty minutes, or less, if they completed the paper before the signal to stop.

The second method was adopted because the fourth page was often incomplete on the test papers from grade ten. Here the examiner read the first page aloud, pausing for the pupils to fill in the blanks, and answering questions regarding the directions. A signal was then given to begin page two. After ten minutes the direction was given to turn to page four. Thus the time was equally divided between pages two and four.

Results. The results of the first trial of The Berkshire Word Test in grades 7B through 12A are given in the two accompanying tables, X and XI, which present in grade score frequencies the records of pages two and four respectively. For each page there is a rise in percentage of successes through the group but the increase is so gradual that one list of fifty words could not be trusted to discriminate sharply between grades.

For page two the median score for 1289 cases was 21 with a median deviation of 6: the average score was 22 with a standard deviation of 8.9. The corresponding figures for page four were: median, 18 with a median deviation of 5; average, 19 with a standard deviation of 7.4. As in 1922, the test on page four rated somewhat below the one on page two, there being more scores of nine per cent or less, and fewer of thirty per cent or over. The respective coefficients of variability for the two pages are 27 and 34.

TABLE X

Showing Number of Cases, by School Grades, at Each Interval of Score

Page 2 Scores							Grades						
	7B	7A	8B	8A	9B	9A	10B	10A	11B	11A	12B	12A	Tot.
40–44							1		3	4	2	4	14
35–39		1		1		1	8	7	5	8	8	11	50
30–34			1	1		1	25	12	22	11	13	25	111
25–29		3		5	2	11	70	26	39	27	34	25	242
20–24	4	8	11	14	6	16	91	32	44	15	31	20	292
15–19	9	27	19	28	20	18	106	29	29	9	12	10	316
10–14	21	36	18	21	14	12	40	10	4	2	3	2	183
5–9	16	20	8	7	7	4	11	2		1			76
0–4	2	1	1	1									5
Total	52	96	58	77	50	63	352	118	146	77	103	97	1289

Median	21	Average of distribution	22
Median deviation	6	Standard deviation	8.9
	Probable error of score	3.06	

TABLE XI

Showing Number of Cases, by School Grades, at Each Interval of Score

Page 4 Scores							Grades						
	7B	7A	8B	8A	9B	9A	10B	10A	11B	11A	12B	12A	Tot.
40–44												1	1
35–39						1	5	1	4	3	1	8	23
30–34		1		1		2	9	8	13	9	10	10	63
25–29		1	3	5	2	9	43	14	24	15	19	26	161
20–24	5	8	4	11	4	16	81	29	43	20	24	27	272
15–19	8	18	18	24	15	15	110	32	38	16	31	15	340
10–14	19	36	28	24	22	12	76	28	15	9	15	8	292
5–9	19	28	4	12	7	7	24	5	6	5	3	2	122
0–4	1	4	1			1	4	1	3				15
Total	52	96	58	77	50	63	352	118	146	77	103	97	1289

Median	18	Average of distribution	19
Median deviation	5	Standard deviation	7.4
	Probable error of score	2.55	

The relative difficulty of the three pages of the test is indicated by the averages of grades 10B through 12A of the public schools and one hundred pupils, grades 9 through 12 of Miss Hall's School, which are presented in the next table. The grade figures here are based on the results obtained by the first method of testing described on page—. Since the method of grading in Miss Hall's School is unlike that of the public schools, the average of that group is given separately. Moreover an attempt was made with them to discover what influ-

ence an order to guess would have; therefore, a different method of testing was employed.

TABLE XII
Berkshire Word Test
Average Number of Words Right

Grade	No. Cases	Page 2 Av.	σ	Page 3 Av.	σ	Page 4 Av.	σ
12A	97	27.69	6.6	23.5	5.9	23.93	7.2
12B	103	25.7	6.04	23.58	5.02	20.6	6.2
11A	100	26.2	7.0	23.2	5.8	21.46	7.5
11B	146	24.5	6.1	22.6	5.7	20.8	6.9
10A	73	22.7	6.3	20.6	5.1	19.0	6.6
10B	145	21.6	6.5	19.8	5.4	18.51	6.5
12A, 12B, 11A	212	26.9	5.9			23.2	8.4
M.H.S.	100	32.2	6.3	26.36	6.0	29.5	6.6*

* Directions to guess.

From these figures it appears that page two is simpler than either page three or page four. Pages three and four seem to be equivalent for 12A pupils, but less advanced students failed to make as good a score on page four as on three. Investigation disclosed the fact that the younger pupils, who often failed to apportion their time wisely, were unable to complete page four. In the tenth grade this difficulty was most frequently seen.

Grade 12B often scored less than 11A. Whether this was due to limitations of the test or of the group is a question. But since the same situation arose with page three, which is a standard word list, and with the academic averages for the same semester in which the test was given, it seems safe to conclude that 12B developed more slowly than 11A; therefore, the figures from this test are not such reliable norms for pupils of that grade. Careful study of grade averages indicates that the score of 11A is high in the test, thus increasing the difference between the two grades.

Further trial of the tests was made in the second semester in order (1) to improve, if possible, the scores of page four, (2) to study the effect of controlling more carefully the time spent on each page, and (3) to see the effect of an order to guess.

In February, 1927, pages two and four were repeated in Miss Hall's School with a change of method. The instructions given by the examiner in December rendered the test a meas-

ure of silent reading ability as well as of vocabulary. To obviate this complication, the examiner read aloud the directions on the first page. When the task seemed to be thoroughly understood, the signal was given to begin page two. At the end of ten minutes all were directed to turn to page four. At this time was given the additional order to guess at unfamiliar words.

The averages obtained in both tests were greater than on the previous trial; the increase for page two was 3.7; that for page four, 3.4. The standard deviation for page two was reduced from 6.92 to 6.32, while that for page four remained the same, 6.62 on both occasions.

TABLE XIII
Average Scores and Standard Deviations of Miss Hall's School in Two Tests

Page	December		February		Gain
	Av.	*St. Dev.*	*Av.*	*St. Dev.*	
2	28.5	6.92	32.2	6.32	3.7
4	26.1	6.62	29.5	6.62	3.4

The gains made on the two pages in the retesting are so nearly equal that the influence of guessing is negligible. It is probable that the improvement is due to practice and the different method of giving instructions, facts which apply equally to both pages.

Near the close of the second semester the second method of directing the test was tried in grade 12A with seventy-six pupils who took the test in the first half of the year. These pupils showed improvement on both pages; but the greater gain is on page four, which suggests that the time control was needed, as page four was improved with no sacrifice of page two.

$$\text{Page two } 27.97 \pm 6.16 \text{ and } 30.0 \pm 6.26$$
$$\text{Page four } 22.13 \pm 6.6 \text{ and } 26.5 \pm 6.26$$

Besides the retests, new investigations were started with the changed instructions in grades 7B through 10B. In Table XIV the average scores with the standard deviations appear by semesters. Since the later method of administering the test affected the scores in the middle range very slightly, avrages of all grades studied are included in the same table.

These averages are based on the total scores made on pages two and four. The increase in score between succeeding semes-

TABLE XIV
General Averages

Grade	7B	7A	8B	8A	9B	9A	10B	10A
Cases	51	96	58	77	50	63	145	73
Av.	23.6	26.2	29.8	31.9	29.1	37.4	40.1	41.7
σ	9.5	10.9	10.1	11.6	9.5	13.2	13.0	12.9

Grade	11B	11A	12B	12A	M.H.S.	Distribution
Cases	72	76	74	62	100	997
Av.	45.3	48.6	46.3	51.6	54.8	41
σ	13.0	14.7	12.4	13.8	13.7	12.8

ters is continuous, the difference, in the case of all grades except 9B and 12B, varying between 1.6 and 3.6, with an average gain of 2.1 words.

No marked lowering of the average score resulted from limiting the time and requiring a double response. On comparing the 8A averages of 1922 and 1926 a difference of 1.46 words is found on page two and of 1.9 words on page four. This difference of one third of the standard deviation is far outweighed by the increased efficiency of the test.

The irregularity in the grades mentioned above, 9B and 12B, is due to accidents in school grading. These two classes happened to be a poor sampling of the school population. To establish the fact that this apparent failure of the word list was due to the groups rather than to the test, teachers were consulted and the score of these same pupils on the Terman list on page three of the test was compared with the Berkshire Word Test averages. The 8A average for page three was 16.14 ± 5.2; the 9B average for the same page was 16.15 ± 4.5. In grade 11A the average for page three was 23.22 ± 5.8 and that of 12B was 23.58 ± 5.02. This standard scale also fails to discriminate between grades at these points. All evidence, both teachers' statements and test results, indicated that the two grades in question were less successful in exercises of this character than other school groups.

Because of differences in school organization the results of private school pupils could not be combined by grades with the public school records, but age-score averages of all subjects are presented in Table XV.

TABLE XV
Showing Age-Score Averages of 1441 Cases in Grades 7-12

Age	Number of Cases	Average Score	Standard Deviation	Coefficient of Variation
10	2	14.5	....	....
11	23	15.0	3.1	20.6
12	76	13.7	5.65	41.6
13	141	16.2	5.25	32.4
14	198	18.6	7.0	37.6
15	312	20.8	6.35	30.5
16	325	21.5	6.15	28.6
17	222	23.4	6.85	29.3
18	112	23.6	7.05	29.9
19	24	20.8	6.44	31.0
20	6	23.7	6.65	28.1

The curve of averages presented in this table rises continuously from twelve through seventeen years, but is flattened at eighteen; at ten and eleven it is relatively too high. The ten and eleven year groups are too small to be representative of the age. These subjects are mentally superior, as their scores and school grade both indicate. The groups beyond eighteen years are also small because relatively few pupils over that age remain in the public schools.

Summary of test results.

In the period of research described in this chapter 1441 pupils were tested. Of this total 185 subjects took five tests of fifty units each and the remaining 1256 took three tests of fifty units each. Thus the tables given below are based on results obtained from 234,650 definitions.

The public school scores have been summarized in two tables which indicate the per cent of cases reaching or exceeding each point of score from five to forty. In Table XVI 1345 are grouped by age in years; in Table XVII the classification of 1347 is based on school grading, the two semesters of each grade being averaged together. In each table the raw scores appear in the first column, while in the last column are given percentages of the whole distribution reaching or exceeding each score. Two ten-year cases which appear in the grade scale are omitted from the age scale, this making the total number in table XVI 1345 instead of 1347.

The median of the whole distribution of 1347 is 19.66 ± 4.5. The average of the distribution is 19.69 ± 6.81 making 34.6 the coefficient of variation.

TABLE XVI

Vocabulary-Age Scale Based on List of Fifty Words Showing Per Cent of Each Age Reaching or Exceeding Each Point of Raw Score

Raw Score	Age in Years									Per Cent of 1345 Cases
	12	13	14	15	16	17	18	19	20	
40				1	0	1	1			0.2
39				1	0	2	2			
38				1	0	2	3			
37				2	1	3	4			
36				2	1	4	5			
35			1	2	2	5	6			1
34			2	3	3	7	8	1	3	2
33			3	4	5	9	11	2	6	4
32			4	6	7	11	14	4	9	6
31			5	8	9	14	17	6	13	8
30	1	2	6	10	11	17	20	8	17	10
29	1	3	8	13	14	21	23	13	23	13
28	2	4	11	16	17	25	26	18	29	16
27	3	5	14	19	20	30	30	23	36	19
26	4	6	17	22	24	35	34	28	43	23
25	5	7	20	25	28	40	38	33	50	27
24	8	11	24	31	35	46	45	37	57	32
23	10	14	28	37	42	53	52	41	64	37
22	12	17	32	43	49	59	59	45	71	42
21	14	20	36	49	57	65	66	50	77	47
20	16	23	40	55	64	71	73	54	83	53
19	20	29	46	62	69	76	78	59	83	58
18	24	35	53	68	74	81	82	64	83	63
17	28	42	60	74	79	86	86	69	83	68
16	32	49	67	80	83	90	90	70	83	72
15	35	56	73	86	87	94	94	79	83	76
14	44	63	77	89	90	95	95	84	87	80
13	52	70	81	92	93	96	96	88	91	84
12	60	76	85	94	95	97	97	92	94	88
11	69	82	89	96	97	98	98	96	97	92
10	78	88	93	98	99	99	99	100	100	95
9	82	91	95	98	99	99	99			96
8	86	94	97	99	99	99	99			97
7	90	96	98	99	100	100	100			98
6	95	98	99	100						99
5	100	100	100							100

A VOCABULARY INFORMATION TEST

TABLE XVII

Vocabulary-Grade Scale Based on List of Fifty Words Showing Per Cent
of Each Grade Reaching or Exceeding Each Point of Raw Score

Raw Score	Grade						Per Cent of 1347 Cases
	7	*8*	*9*	*10*	*11*	*12*	
39							
38							
37							
36							
35			1	1	3	4	2
34			1	2	4	6	3
33			2	3	6	9	4
32			3	4	8	12	5
31			4	5	10	15	6
30	1	1	5	6	12	18	7
29	1	1	6	8	16	23	9
28	1	2	7	10	20	29	11
27	2	3	8	12	25	35	14
26	2	4	9	15	30	41	17
25	2	5	10	18	35	47	20
24	3	8	16	24	42	54	25
23	4	·11	22	30	50	61	30
22	5	14	28	36	58	67	35
21	7	17	34	43	66	73	40
20	9	20	40	50	74	79	46
19	13	28	47	57	79	83	52
18	17	36	54	64	83	87	58
17	21	44	61	72	87	91	64
16	25	52	68	78	91	94	69
15	30	60	75	84	95	97	74
14	39	67	79	87	96	98	78
13	48	74	83	90	97	98	82
12	58	81	87	93	98	99	86
11	67	87	91	96	99	100	90
10	76	93	95	98	100	100	94
9	81	95	96	98	100		95
8	86	97	97	99	100		96
7	91	98	98	99	100		98
6	96	99	99	99	100		99
5	100	100	100	100			100

The figures of page three are not treated separately in this summary of the test results because it is not the purpose of this research to establish norms on Terman's list, a psychometric instrument of well-known value. This list was inserted between the pages of the present test with the permission of the pubishers, Houghton Mifflin Company, so that it might be employed as a criterion in studying the difficulty of the Berkshire Word Test.

The vocabulary-age medians in table XVI rise from age twelve through seventeen. From eighteen through twenty, the line of medians is irregular because of the small number of these cases present in the public school.

The vocabulary-grade medians in table XVII rise from grade seven through grade twelve. The most rapid increase in vocabuary medians appears between grades seven and eight, after which the medians are marked by a steady rise of two points each year representing a yearly vocabulary gain of four per cent.

There is great overlapping of both grades and ages, which can be more plainly seen in Table XVIII where the mode age of each grade is placed at the bottom of the grade column, thus making possible easy comparison of the grade and age ratings.

TABLE XVIII

Showing Per Cent of Each Age and Grade Reaching or Exceeding Each Grade Median

Grade Medians	Grade					
	7	8	9	10	11	12
25	2	5	10	18	35	47
23	4	11	22	30	50	61
20	9	20	40	50	74	79
18	17	36	54	64	83	87
16	25	52	68	78	91	94
13	48	74	83	90	97	98
Mode age	12	13	14	15	16	17

CHAPTER VI

BERKSHIRE WORD TEST COMPARED WITH OTHER MEASURES

The test results recorded in the preceding chapter showed that the word lists descriminated ages twelve through seventeen and grades 7 through 12. The next task was to compare these results with those of reliable standard tests in order to discover the degree of correspondence and also to determine the relative difficulty of the Berkshire Word Test. Accordingly scores in this test have been correlated with results of other tests and school records.

The Stanford Revision of the Binet-Simon Tests

The age score medians of the Berkshire Word Test and of the Terman list are presented below.

TABLE XIX

Showing the Age Standards in Two Tests Based on Fifty Word Lists

Age	10	12	14	16
Terman Norm.	15	20	25	33
B. W. T. Median	..	13	18	23

Evidently the Berkshire Word Test is more difficult than the Terman vocabulary in the age range for which standards have been computed, since the median at each age is less than Terman's standard for the next lower age group.

Thorndike Word List

Both the Terman list and the Berkshire Word Test give a more accurate representation of the word knowledge of pupils in grades 7 through 12 than the Thorndike Word List for the reason that they are both based on a random sampling of a more varied list of words. Moreover the Thorndike list presents such words as pupils of the first eight grades should be taught, while the aim of the Berkshire Word Test, as of the Terman list, is to find out what words children do know rather than what they should know. The Thorndike lists are particularly selected to aid in teaching; the Berkshire list is designed

to aid in job analysis. Again the clinical lists are intended for use with cases varying in intelligence from an eight-year level to that of the superior adult while the Thorndike list is adapted particularly to the language work of children ranging from six to fourteen years, the population of the first eight grades. Many words which are familiar to high school pupils are missing from this list. Following are a few of the words not found in the Thorndike Word List, all of which appeared on the papers of pupils in grades 9 through 12: constellation, exit, geology, gnome, gristle, sticky, windpipe.

In order then to test fairly the knowledge of pupils above grade 8 it is necessary to employ a sampling of a varied list suited to the rapidly increasing vocabulary of these grades. Since the reliability of such a test increases "directly as the square root of the number of words used"* the Berkshire Word Test has been selected from a longer list than the Terman test, thus furnishing a selection of greater variety with more words of approximately equal difficulty.

That selecting words from the Collegiate Dictionary resulted in a more difficult list is shown by a comparison of the Berkshire Word Test with the Thorndike Word List. Of the stimulus words given on pages two and four, forty per cent are not found in The Thorndike Word Book; of those on page three, the Terman list, thirty-two per cent do not appear in that same list. Of six hundred response words in the Berkshire Word Test, forty-two, or seven per cent, are not found in the Word Book; but since all were definitions used by pupils in grades 9 through 12, it is fair to include them in the test. Thirteen of these response words, about two per cent, are correct definitions. Disregarding, then, all incorrect response words, there remain sixty-nine essential words which are not included in Thorndike's list. A part of these words seem no less familiar to high school pupils than many that are in the Word Book, but the fact that fifty-seven of them occur in the last half of their respective lists is significant of the greater difficulty of the Berkshire Word Test.

Both the Terman list and the words of the Berkshire Word Test have become better adapted to measuring in the high school grades by changing the form of the responses from oral definitions, which were scored very liberally, to written re-

---

*Sixteen Spelling Scales, Standardized in Sentences for Secondary Schools, page 7.

sponses of the multiple choice type which require more precise definitions.

The National Intelligence Test

Correlation coefficients resulting from the comparison of vocabulary scores with those obtained in the National Intelligence Test are presented in Table XX. These, like all correlations reported in this investigation, have been secured with the help of the Otis Correlation Chart (28), the formula of which is equivalent to the formula:

$$r = \frac{\Sigma x^2 + \Sigma y^2 - \Sigma v^2}{2\sqrt{\Sigma x^2\, \Sigma y^2}}, \quad k = \sqrt{1-r^2}, \quad P\,E = .67\,\frac{1-r^2}{\sqrt{n}}$$

TABLE XX
Correlations of B. W. T. with the N. I. T.

Grade	No. of Cases	Page 2			Page 4		
		r	P E	k	r	P E	k
12A	80	.53	.054	.85	.62	.046	.78
12B	71	.54	.056	.84	.33	.070	.94
11A	77	.41	.064	.91	.50	.057	.86

		Pages 2 + 4			Page 3		
11B	72	.43	.064	.90	.48	.061	.88
10A	73	.44	.062	.90	.33	.067	.94
10B	145	.53	.040	.85	.52	.040	.85

Since the scores which were the basis of the foregoing table resulted from tests that were separated by a period of four years, a close correlation could hardly be expected. It is interesting, however, to note that all coefficients based upon these figures are positive, thus indicating a similarity between the tests.

The Stanford Achievement Test

The Stanford Achievement Test was given in 1923 to 132 pupils of grade 6A, who took the Berkshire Word Test in the 10B grade of 1926. The following table shows that the scores of these two tests correlate more closely than did the scores of the Berkshire Word Test and the National Intelligence Test.

TABLE XXI

Correlations of Berkshire Word Test with the Stanford Achievement Test

| Grade | Cases | Pages 2 + 4 | | | | Page 3 | |
		r	P E	k	r	P E	k
10B	132	.75	.025	.66	.71	.029	.71

The Otis Classification Test

Since 1924 the Otis Classification Test has been given each year to pupils in grade 6A and 7B. The figures used in this report are based on scores collected within the past year.

The four lines of Table XXII give correlations with the scores obtained from the Otis Classification Test as follows: the total score, the mental age, the achievement score, and the classification index. The largest coefficient resulted from the correlation of the classification index with the score on 100 lines of the Berkshire Word Test. This classification index is a figure which represents both native ability and achievement.

TABLE XXII

Correlation of Berkshire Word Test with the Otis Classification Test in Grade 7B Nov., 1926

| Score | Pages 2 + 4 | | | | Page 3 | |
	r	P E	k	r	P E	k
O. Cl. Total	.45	.074	.89			
O. Cl. MA	.54	.066	.84	.50	.067	.87
O. Cl. Ach.	.68	.050	.57	.64	.056	.77
O. Cl. CI	.74	.042	.67			

Haggerty Reading Examination and Otis Self-Administering Test of Mental Ability

Near the beginning of each school year the Haggerty Reading Examination and the Otis Self-Administering Test of Mental Ability are given to all pupils of Miss Hall's School. These tests have furnished relative ratings that are helpful in classifying new pupils, but in order to distribute the group it is necessary to limit the testing time to twenty minutes. For the past two years, the method of conducting both tests has been the same as that suggested in the Otis Self-Administering Test of Mental Ability, higher form.

Correlations were computed with the total score in these two tests and with the Haggerty Vocabulary alone. It is interesting to compare these two vocabularies which are constructed on the same plan, each list containing fifty lines.

The average scores of 102 pupils in these vocabularies follow: Haggerty, Sigma 3, Form A   41.8 ± 4.6
    B.W.T. page 2   28.7 ± 6.9
    B.W.T. page 4   26.1 ± 6.8

These figures indicate the superiority of the Berkshire Word Test for measuring the upper grades of the high school.

Table XXIII shows the correlations of the tests under consideration.

TABLE XXIII

Correlations of the Berkshire Word Test, Haggerty Reading Examination, and Otis Self-Administering Test of Mental Ability, Form A, Based on Records of 102 Girls in Miss Hall's School

| Test | Page 2 + 4 | | | Page 2 | | | Page 4 | | |
	r	PE	k	r	PE	k	r	PE	k
Hg. Rd. To.	.58	.043	.81						
Hg. Voc.	.79	.025	.61	.70	.030	.71	.70	.030	.71
Otis S-A	.63	.040	.77						

In the form of response employed, the Berkshire Word Test is similar to the Vocabulary Test in the Haggerty Reading Examination with which page two gives a correlation of + .704 ± .03 and page four a correlation of + .700 ± .03 (Table XXIII). The Berkshire Word Test, however, is more useful in grading the advanced pupils as is shown by the fact that in the Haggerty Reading Vocabulary of fifty words, seventy-two per cent of the pupils in Miss Hall's School defined forty or more words correctly while in the Berkshire Word Test only six of the same pupils defined correctly forty words of page two and only four pupils defined as many as forty words on page four.

As might be expected, the correlation is greatest between the two vocabularies. In the total score for the Haggerty Reading Examination, where the correlation is least, comprehension and memory have more weight than with the vocabulary. When the Otis Self-Administering Test of Mental Ability was correlated with the total score of the Haggerty Reading Examination, the coefficient was + .58 ± .043, just

equal to the correlation between the Reading Examination and the sum of pages two and four of the Berkshire Word Test.

These correlations emphasize some facts about the Berkshire Word Test: 1. As a measuring scale it resembles the Haggerty Vocabulary; but, because of its difficulty, it discriminates better in the upper grades.

2. This test resembles the Haggerty Reading Examination no more than does the Otis "Test of Mental Ability."

3. This test correlates more closely with the "Test of Mental Ability" than with the complete Reading Examination.

School Marks

Although the group tests described above need to be combined with the judgement of teachers when pupils are classified, still they furnish valuable assistance in that work. These test results have in several instances been correlated with school marks to see what relation exists between them.

At the close of the twelfth grade in the Pittsfield public schools each pupil's "general average" is computed, a figure which represents all the work of three years. For pupils below that grade the most comprehensive mark is a semester average. For the following computation the average corresponding most nearly to the date of the test was selected. Correlations of the test with these two averages are given in Table XXIV.

TABLE XXIV

Correlations of Berkshire Word Test with School Marks of 245 Pupils in Grades 11A through 12A

Grade	No. of Cases	School Average	Test Page	r	P E	k
12A	95	General	2 + 2	.49	.061	.70
12A	67	General	4	.57	.055	.75
12A	62	General	2 + 4	.57	.057	.82
12B	74	Semester	2 + 4	.41	.064	.91
11A	76	Semester	2 + 4	.38	.066	.93

From Table XXIV it appears that there is greater agreement when correlations are based on general averages than when semester averages are used. This may be due to the comprehensiveness of the average or to the greater maturity of the group, as well as the character of the work. No correlation in this table is high enough to serve as a basis for pre-

dicting individual school grades, though the test is useful in grade surveys.

In Miss Hall's School, where promotion is by subjects, the only mark available was an average of the marks earned in academic courses during the same semester in which the tests were given. Below are correlations of three tests with this average and with examination marks in separate subjects.

TABLE XXV

Correlation of Three Word Tests with School Marks of 100 Girls in Miss Hall's School in Grades 9B Through 12A

Test	No. of Cases	Semester Average		
		r	P E	k
Otis S-A. Form A	100	.45	.054	.89
Hag. Read. Σ3, Form A	100	.44	.054	.90
B. W. T. 100 lines	100	.37	.057	.93
B. W. T. 50 lines	95	.38	.056	.92
B. W. T. 250 lines	94	.46	.054	.89

In this table the largest coefficient was obtained when the scores of five word lists were averaged, but even then the result was only .46. Five word lists would have the same predictive value in this group as one Haggerty Reading Examination or one Otis Self-Administering Test. The correlation was increased but slightly by lengthening the word list to 250 words.

A comparison of the vocabulary tests with examination marks in mathematics, French, and Latin follows:

TABLE XXVI

Correlation of Berkshire Word Test with School Marks in Mathematics, French and Latin

Subject	No. of Cases	Berkshire Word Test	r	P E	k
Mathematics Exam.	68	Percentile rating	.26	.026	.57
French Examination	88	P. Rating	.58	.048	.81
French Examination	93	Av. of 250 lines	.45	.056	.89
Latin Examination	48	P. rating	.56	.066	.83

The greatest correlation was found between the French examination marks and the percentile ratings in the Berkshire Word Test which was based on the average of the scores obtained in five lists. The correlation of these same ratings with

the Latin examination marks resulted in a coefficient of .56; but this group was much smaller than the French group, since Latin is not required throughout the course.

Summary

The comparisons of the Berkshire Word Test with other test results indicate that the test:

1. measures in varying degrees the traits which are measured by the standard tests employed;

2. discriminates between ages twelve and seventeen as sharply as the Terman Vocabulary;

3. correlates positively with two measures of native ability, the National Intelligence Test and the Otis Self-Administering;

4. correlates positively with two measures of acquired ability, The Stanford Achievement Test and the Haggerty Reading Examination;

5. correlates positively with all measures obtained from the Otis Classification Test, but most closely with the "Classification Index";

6. correlates positively with school marks and when the average of five lists was used the correlation was greater than that of the Otis Self-Administering or the Haggerty Reading Examination with the same marks.

The correlations obtained thus far, though positive, are not of such order as to suggest that the test duplicates any test with which it has been compared.

CHAPTER VII

RELIABILITY OF THE BERKSHIRE TEST

Reliability coefficients of the Berkshire Word Test have been calculated in two ways, by repetition of the same form and by correlation of similar forms. Since test forms and the groups tested have been described earlier, the tables of this chapter will present only the data relating to the self-correlations and intercorrelations of the test forms. In these calculations, page three has been treated as one part of the Berkshire Word Test; since, as a multiple choice test, it is more nearly equivalent to this than to the original test from which it was developed, the Terman vocabulary (38), published by Houghton Mifflin Company.

With each correlation coefficient and probable error, these tables give the corresponding index of reliability, the square root of the correlation coefficient, which Ruch and Stoddard say is "the limit of correlation of anything with the test under consideration" (55).

Table XXVII records the self-correlations of the Berkshire Word Test based on data obtained in 1926 and 1927 when the same tests were used twice in Miss Hall's School.

TABLE XXVII
Self-Correlations of the Berkshire Word Test

Test Page	Group Tested	Number of Cases	r	PE	$\sqrt{r}$
2	11B-12A	76	.77	.031	.88
2	M. H. S.	98	.78	.027	.88
4	11B-12A	76	.63	.046	.79
4	M. H. S.	98	.87	.017	.93
2 + 4	12B-12A	67	.88	.017	.94
2 + 4	M. H. S.	97	.87	.016	.93
2 + 4	M. H. S. + 12A	173	.85	.014	.92

In this table the reliability coefficient of page two is rather small, but approximately equal in the two groups tested; that of page four is low in the public school and fairly high in Miss Hall's School. Were the same thing true of page two this might have been explained by the fact of the greater range of grades in the latter group; but, as the figures stand, a different cause must be sought. Probably several subjects who in

the first test left page four incomplete improved their scores in the retest, because the time was evenly divided between pages two and four.

The reliability is fairly high in all groups when the length of the test is doubled, showing that a list of one hundred definitions would furnish a "fairly adequate" (55) individual measurement.

In Table XXVIII are the coefficients resulting from the correlation of similar test forms.

TABLE XXVIII
Correlation of Three Similar Forms of the Berkshire Word Test

Grades	No. of Cases	Page	r	3 PE	$\sqrt{r}$	r	4 PE	$\sqrt{r}$
7B-12A	1306	2	.76	.007	.89	.74	.009	.86
7B-12A	1306	3				.75	.008	.87
12	67	2				.76	.034	.87
M. H. S.	97	2 + 2		with 4 + 4		.84	.020	.92

For the group of 1,306 reported in the table the low reliability coefficients show that the test is not reliable throughout the distribution. In the lower grades tested, the averages pointed out the difficulty of the tests, especially that of page four. Also the guessing of the pupils caused undue variability in their scores. In small groups of the more advanced pupils the reliability is equal to that in the larger heterogeneous group. The highest coefficient obtained, that in Miss Hall's School, (.84 ± .02) was based on the average of two trials each on pages two and four.

The intercorrelations of the test scores of grades 10B through 12A are given in Table XXIX. This table is based on the results of tests given by "method II" which is fully described in the Examiner's Guide, Appendix, page 67. It is the method employed in the Otis Self-Administering Test.

The correlation coefficients in the following table are small because of the small range of ability included in the group.

In Table XXX are the intercorrelations of the same test forms which resulted from the scores of 100 girls in Miss Hall's School.

The average of the coefficients of the 766 records included in Tables XXIX and XXX is .73 ± .014 when pages two and

TABLE XXIX
Intercorrelations of Pages 2, 3, and 4 in Grades 10B-12A in 1926.
(Old Method)

| Grade | No. of Cases | Page | r | 3 | | r | 4 | |
				P E	$\sqrt{r}$		P E	$\sqrt{r}$
12A	97	2	.72	.032	.85	.80	.025	.89
		3				.74	.030	.86
12B	103	2	.68	.035	.82	.56	.045	.75
		3				.55	.046	.74
11A	100	2	.78	.026	.88	.67	.037	.82
		3				.63	.040	.79
11B	146	2	.56	.038	.75	.55	.038	.74
		3				.51	.041	.71
10A	73	2	.67	.034	.82	.76	.025	.87
		3				.67	.034	.82
10B	145	2	.74	.020	.86	.66	.026	.81
		3				.57	.038	.76

TABLE XXX
Reliability Coefficients Based on Scores of 100 Girls in Miss Hall's School
Grades 9B-12A

| Page | r | 2 | $\sqrt{r}$ | r | 4 | $\sqrt{r}$ |
		P E			P E	
2	.78	.027	.88	.77	.027	.88
3	.74	.030	.86	.73	.031	.85
4				.87	.017	.93

four are compared. Such a coefficient Otis (28) suggests is as accurate as one coefficient on a large number of cases (page 260).

If .78 represents truly the degree of resemblance between pages two and four, it would be necessary to use only three such lists to secure a reliability coefficient of .91 according to Brown's formula. Or according to Garrett's formula (page 273) the "index of reliability" would be the square root of .78 which is .88, "the highest self-correlation which can be obtained (except by chance) with this test in this form."

When corrected for attenuation the correlation of pages two and four is .94. The coefficients show that the different forms of the test are of approximately equal reliability in this group. According to the interpretation of such coefficients given by Ruch and Stoddard (55), this form would be "fairly adequate for individual measurement."

In these two tables no index is less than .77 and one is .93. In any grade the average score in three lists would give a re-

liable rating of an individual pupil. The score on one list would be very suggestive in classification and sectioning classes.

Reliability of testing methods

In order to make a more thorough study of the changed instructions, a trial test was made in grade 7B of the Pomeroy School where fifty-one complete records were secured. This grade is quite heterogeneous, including pupils from several outlying districts. The average age is twelve years, six months. The principal, who was interested to try the test, assigned the work of administering it to the English teachers.

These pupils took the "Otis Classification Test" either in November, 1926, or if they were repeating the grade, in June, 1926, thus making possible correlation with a standard test given within the year. Correlation coefficients have been calculated with page 3 as a criterion, the Terman list, and with each part of the Classification Test. The Berkshire Word Test correlates a little more closely with the Achievement Test (.64 $\pm$ .056) than with the Mental Age figure (.50 $\pm$ .042) which is obtained from a comparison of page 2 + 4 with the Classification Index. It is interesting to note on the line of averages that 7B fills in the place that was forecast for it by the direction of the line of averages of grades 10-12 (Table XII, page 35).

The next application of the modified method was made in 10B among the children who in December found the test time too short. Except for a few who were repeating the grade, a wholly new group was examined in April, 1927. In making the correlations the few records of those tested in December were excluded in order to see what effect the change of method produced on the records of children unfamiliar with the lists.

Tests were soon given to all the junior high school pupils in the Pomeroy School. The results of these tests given with the later method are recorded in Table XXXI. For the exact instructions employed, see Examiner's Guide, Appendix, page 66.

The reliability index of the whole 531 cases is an actual correlation obtained by the use of the Otis Correlation Chart. A comparison of the results from the two testing methods is afforded when the correlation coefficients of 10B scores in Table XXIX are compared with those of 10B scores in XXXI. Equalizing the time increases the reliability according to two

TABLE XXXI

Intercorrelations of Three Pages of the Word Test in Grades 7B-10B.
(New Method)

Grade	No. of Cases	Page	3			4		
			r	P E	$\sqrt{r}$	r	P E	$\sqrt{r}$
10B	135	2	.68	.031	.82	.72	.027	.85
		4	.71	.029	.84			
9A	63	2	.78	.040	.88	.76	.036	.87
		4	.76	.036	.87			
9B	50	2	.59	.062	.77	.62	.058	.79
		4	.71	.047	.84			
8A	77	2	.69	.040	.83	.81	.026	.90
		4	.68	.041	.82			
8B	58	2	.68	.048	.82	.74	.040	.86
		4	.62	.054	.79			
7A	96	2	.71	.034	.84	.70	.035	.84
		4	.71	.034	.84			
7B	52	2	.63	.056	.79	.59	.060	.77
		4	.59	.060	.77			
Total	531	2	.75	.012	.87	.77	.012	.88
		4	.77	.012	.88			

of these intercorrelations. The coefficient of pages two and three is .74 for the December test and .68 for the April group, which suggests that the second group was less successful in adapting itself to the task at the beginning. In the other two correlations a distinct gain is manifest, the correlations for the second group being .74 and .71 as compared with .66 and .57 in the earlier test. Evidently the low scores on the last page were due not only to greater difficulty of the test but also to poor distribution of the time. This experiment shows clearly that the last part of the test suffered from lack of time as well as from inability of the subjects to perform it. In standardizing the test forms it is important to allow a fair amount of time for the completion of each list in order to discover its comparative difficulty.

When "method I," the new method, is used the reliability indices are higher in grade 7B than any that were obtained in grade 11B by "method II." From these results it seems probable that in all grades reliability would be increased by using "method I." Both methods are fully described in the "Examiner's Guide," Appendix, page 66.

CHAPTER VIII

Practical Application of the Test

Next a study was made of the prognostic value of the Berkshire Word Test. In order to render the measures comparable, the school averages of the pupils in Miss Hall's School during the first semester of 1926 and 1927 were transmuted into Berkshire Word Test scores by means of an Otis Percentile Graph (28). By comparison of the individual academic averages with this scale of corresponding values, each pupil's average for April, 1927, was predicted.

Estimated scores were then correlated with the April marks, giving .47 ± .052. The average of the actual marks was 80.67 ± 3.82. Of the ninety-eight cases there were fifty-eight whose predicted marks were within one standard deviation of their true average. In ninety-three cases the difference between true and estimated marks is not over two standard deviations. Since a standard deviation is only four per cent, a difference of two deviations or eight per cent would not cause great change in a pupil's school rating.

Though the correlation coefficient is small, the correspondence between test scores and school marks in the middle range

TABLE XXXII

Correspondence between Berkshire Word Test Scores, Pages Two and Four, and School Marks Based on Marks of 312 High School Students

B. W. T. Pages 2 & 4	M. H. S.	Marks of One Semester 12A	12B	11A	12A, 12B, 11A
85-89	.	95			
80-84	90				
75-79		90		95	100
70-74			95	90	95
65-69	85	85	90		90
60-64				85	85
55-59	80	80	85		80
50-54			80	80	
45-49					75
40-44		75	75	75	
35-39	75		70		70
30-34		70	65	70	65
25-29					60
20-24	70	65	60	65	
15-19			55		55
Acd. Av.	81 ± 3	80 ± 5	78 ± 4.5	78.5 ± 5.5	78 ± 5
B. W.T.	55 ± 9.5	51 ± 9	48 ± 7.8	48 ± 8.8	50 ± 8.75

of the scale is close enough to give the test considerable prognostic value in that range.

The average of this vocabulary score is as valid as one record in the Haggerty Reading Examination or one in the Otis Self-Administering Test.

The correspondence between scores in the Berkshire Word Test and school marks was computed according to the method which Otis (28) describes on pages 106 and 107. Table XXXII shows the resulting figures. Since the distributions compared were linear, the tables were made by reading from percentile graphs corresponding scores found on the same percentile line; that is, scores of equal percentile value in the given distributions. Academic averages for one semester were plotted against the total score of pages two and four in the Berkshire Word Test. According to this table, vocabulary scores of less than 30 occur with school marks below 70 per cent. Vocabulary scores above 50 occur with school marks above 80 per cent. The vocabulary range of the average pupils of this group is between thirty and fifty words on the Berkshire Word Test scale.

This same correspondence between the scores of one hundred pupils in Miss Hall's School is represented in Figure 1, where the average score for three pages is shown by the horizontal line, the academic average for one semester by the vertical.

Through reference to such a graph equivalent scores in two distributions are quickly obtained.
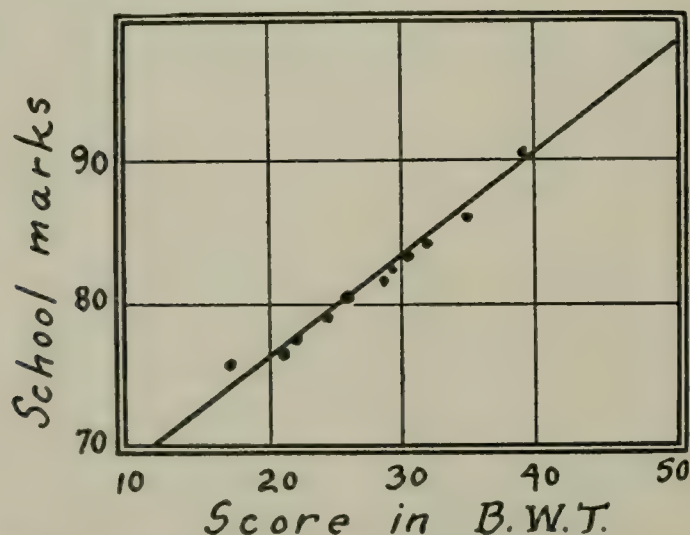
Fig. 1. Showing correspondence between academic average and average of three scores in B.W.T.

A similar relation line resulted when the "general average" of 95 Pittsfield High School graduates was plotted against the average score on two pages of the Berkshire Word Test. This correspondence is shown in Figure two.
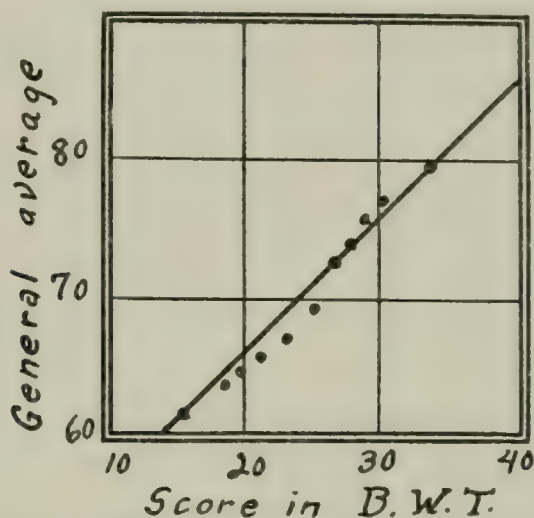
Fig. 2. Showing correspondence between "general average" and score in Berkshire Word Test (two pages).

The "general average" of grade 12A, based as it is on the work of three years, corresponds very closely to the semester average of all grades in Miss Hall's School.

Academic averages predicted from the regression equation were computed by Yule's formula (Yule, page 171). Since the absolute value of the variable derived by this formula was based on only two scores in page two, the error of estimate is so large as to make prediction uncertain.

$$\text{Academic average} = \frac{\text{Test score} - 45.75}{1.236}$$

The error of estimate here is 12.55 on one hundred lines of test, the standard deviation being 13.5 and the correlation coefficient $+.375$. The academic averages predicted by this equation are all smaller than those indicated by the relation lines, but at no point in the distribution is the difference between predicted scores equal to the error of estimate. For example, the school mark corresponding to score twenty is according to the relation line .765, and 69 according to the regression line, showing a difference of 7.5 points.

A multiple correlation obtained by "Formula 15" (28) yielded a coefficient of .50 when the criterion was the academic
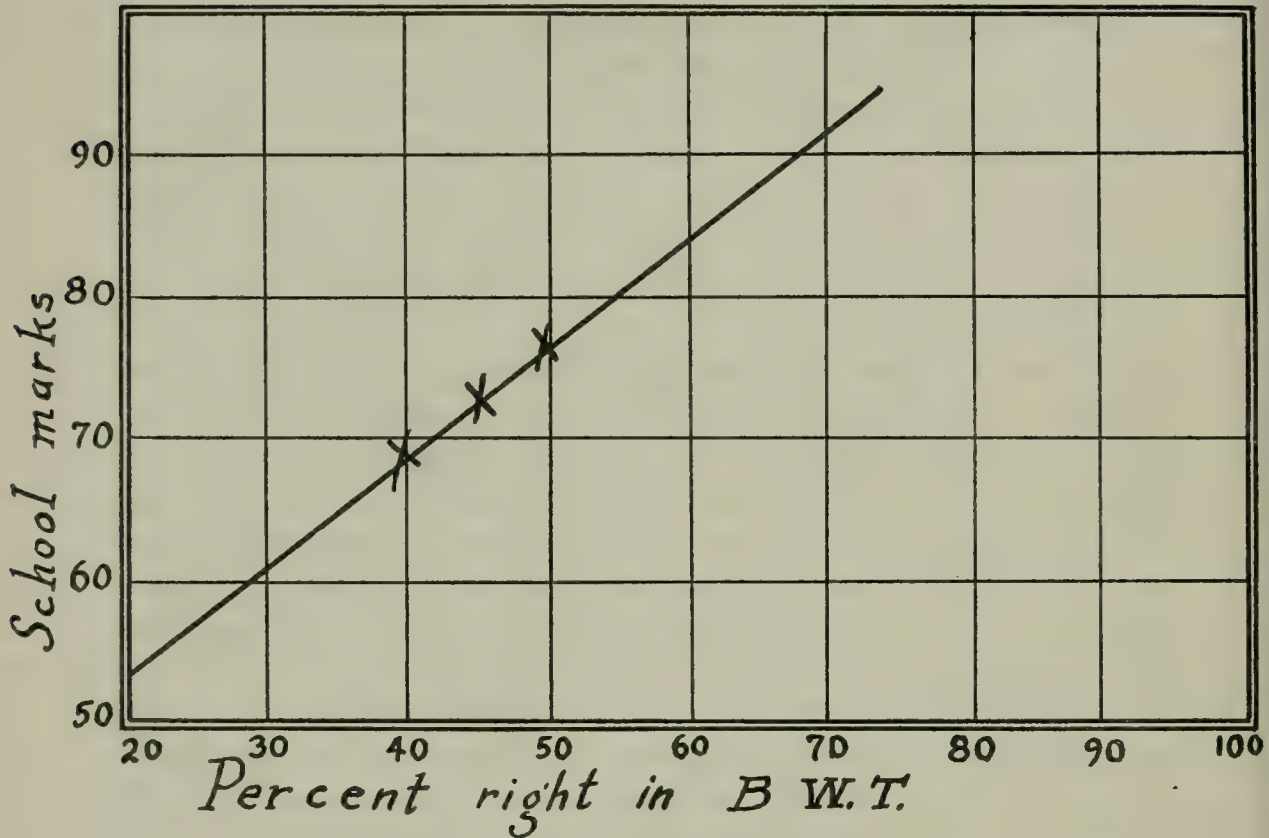
Fig. 3.   Showing the regression of school marks on B. W.T. scores.

average and the scores in the Berkshire Word Test and the Otis Self-Administering were combined as follows:

Weighted score = B.W.T. + .61 Otis S-A.

For convenient interpretation of the scores a scale of percentile ratings by ages has been most useful in schools where promotion is by subjects and grade classification is less regular.   In addition, a scale of percentile ratings by grades is helpful in the public schools.   Such scales, developed from the tables of averages and of grade-score frequency, are given in Chapter IV.   There ratings will be found for six grades, and for ages twelve through twenty.   The age ratings are unreliable above age seventeen and below age twelve because of the smallness of the groups at those levels.

Table XXXIII furnishes a rating scale based on the average score of 1387 cases.   With the help of this scale the score of an individual may be quickly evaluated in terms of the whole population tested.   Similarly a group can be rated by placing its average or median score on this scale.

The low correlations between school grades and test scores agree with the results of other investigators. Arlitt (1), at

TABLE XXXIII
Rating Scale Based on 1387 Records
Standard Deviation of the Distribution 6

Per-centiles	0	0.5	6	30	50	70	94	99.5	100
σ	—3.5	—2.5	—1.5	—0.5	0	+0.5	+1.5	+2.5	+3.5
Score	0	6	12	18	21	24	30	36	42

Bryn Mawr, found a correlation coefficient between IQ and college grades not greater than .298 ± .062. Caldwell (4) who compared adult tests of the Stanford Revision with college grades found the coefficient varying from —.09 to +.69. Chapman (6) concluded that "the differential index" possesses no reliability. Gates (12) after studying the first six grades concluded that group tests were good for rough classification only; they should be called "classification tests" rather than intelligence tests. Haddock (15) found that eighth grade children who had high scores in the Stanford Achievement Test were more successful in the first year of high school. Averill and Mueller (53) obtained coefficients .39 and .46 respectively from Normal School students using Cross English Test, and Inglis Vocabulary. Ohlson (54), who correlated the Terman Group Test with averages in different school subjects, reported coefficients from .24 to .45. This author refers to other investigators whose results are similar.

It is significant that correlations that have been reported are usually positive between school grades and intelligence tests. But the conclusion is inevitable that there are factors entering into school grading which are not found in the tests thus far used. Thorndike (41) asserts that "intelligence is not homogeneous." He found that a test weighted for mechanical ability tends to give different results from one highly verbal in character. If both kinds of tests measure intelligence, then certainly Thorndike's concluson is valid, "intelligence is not homogeneous." But there is still need of an explanation of the fact that marks in school subjects bear so little resemblance to intelligence measures. Either intelligence tests are badly named or school success is surprisingly independent of that trait.

Thoughtful educators are forced to conclude that their problem is very complex. As a clinical thermometer gives important information regarding a patient, so intelligence meas-

ures furnish important data about the mental condition; but neither thermometer nor intelligence test alone would furnish sufficient data for prognosis, nor indeed for diagnosis. Such measures assist in the discovery of exceptional cases, but do not analyze them.

Standard group tests, which aim to eliminate the influence of the examiner, are in their nature different from school tasks which are controlled to a great degree by the instructor and the school situation. On the part of both pupil and teacher school responses are largely subjective; the test responses are largely objective. Herein seems to be one reason for the discrepancy between school marks and intelligence ratings. Test results agree more nearly with school marks when the material of the test samples the content of school courses.

The objection has been raised that achievement tests do not measure native ability. Possibly a more satisfactory name would be classification test or measure of scholastic aptitudes. The name is not so important pedagogically as the fact that the tests distribute pupils according to some ability or combination of abilities essential to success in school work.

Many methods of interpreting test results have been developed, for use until data from reliable and valid group tests can be handled with such precision that prediction therefrom is as accurate as that based on an individual test. But for general use the percentile scale and relation line are very useful.

CHAPTER IX

SUMMARY

Chapter I. The aim of this research has been to devise an efficient vocabulary test that would be a convenient and reliable aid in classifying the pupils of grades seven through twelve.

Chapter II. Since 1908 there has been manifest a growing tendency to apply the results of the earlier vocabulary researches to the solution of psychological and pedagogical problems, to regard the vocabulary index not only as a measure of achievement but also as a measure of mental ability, and to employ these measures for prognosis as well as diagnosis. No longer is there any doubt that the vocabulary is a reliable and valid psychometric instrument. The task of the present is so to improve these tests in form and content that the results obtained through their use may be intelligible to educators and employers.

Chapter III. In the preliminary tests of this series ten fifty-word lists were used, each list being a random sampling of Webster's Collegiate Dictionary of 97,000 words. Using the definitions written by fifty-seven girls as standards, these words were so grouped as to form a scale of uniformly increasing difficulty, with a difference of two per cent between successive steps. From this scale were selected three multiple choice tests, each consisting of fifty stimulus words and two hundred responses. The confusion stimuli presented in these tests were selected from the errors found on the papers of school pupils.

Chapter IV. From 2,502 multiple choice tests that were used in the second trial the following facts were noted:

1. The difficulty of a test line is dependent both on the difficulty of its stimulus word and on its response words.

2. The difficulty of a line may be altered by altering the response words.

3. Each line of a multiple choice test must be treated as a unit.

4. In a multiple choice test of graded difficulty, the relative difficulty of lines must be considered rather than the difficulty of stimulus words employed.

5. The position of the response word in its line influences noticeably the responses marked, especially in some of the difficult lines. Correct answers occur most frequently in the second column and errors in the first column.

6. Some errors are due to careless reading.

7. A period of ten minutes has been found to allow time for a pupil to mark all responses with which he is familiar, in a list of 50 "lines."

Changes suggested by the above observations were incorporated in the third form of the test.

Chapter V. In the third series of tests a folder was used, containing three equivalent word lists of which two, the first and last, were constructed by the author, the other being an adaption of one of Terman's word lists. Two test methods were developed, the first of which left the subject to read the instructions and apportion his time, the second provided that the examiner read the instructions with the pupils and exercise more control over the time. The scores on the last page of the test were higher when the second method was employed, indicating that pupils in these grades profited when receiving more guidance from the examiner.

The averages of grade scores showed a continuous increase in vocabulary, the average gain based on a list of fifty words being two and one tenth words. This suggests a gain in word knowledge, with each semester from grade 7B through 12A, equivalent to about four per cent of a complete adult vocabulary. The age scale shows a continuous gain in word knowledge from twelve through seventeen. At age eighteen a flattening of this curve occurs which is probably due to the fact that few pupils of that age remain in the secondary school. The medians increase less rapidly than the averages.

Chapter VI. Comparison was made of the Berkshire Word Test with other measuring scales as follows: Terman Vocabulary, Thorndike Word List, National Intelligence Test, Stanford Achievement Test, Otis Classification Test, Haggerty Reading Examination, Otis Self-Administering Test of Mental Ability, Form A, and school marks.

With the standard tests used the coefficients were positive but of such order as to suggest that the Berkshire Word Test, while resembling these, does not duplicate any of them. With school marks the coefficients were positive but not high enough to indicate that this test alone could serve as a basis for pre-

dicting individual school grades, though it is as useful in grade surveys as the Otis Self-Administering Test or the Haggerty Reading Examination.

Chapter VII. Through self-correlation of pages two and four the Berkshire Word Test reliability coefficients for three groups were .93, .94, and .92 respectively. The reliability of page two was .88, that of page four was .79 for a group of seventy-six pupils in grades 11B through 12A and .93 for ninety-eight pupils of Miss Hall's School. Intercorrelations of equivalent forms were not quite so high, which indicates that the forms are not exactly equivalent. Intercorrelations when based on 531 records secured by the later method of conducting the test varied from .75 ± .01 to .77 ± .01. According to Brown's formula, three such lists would give a reliability coefficient of .91. One list is very suggestive for classification and sectioning classes. Three lists would be "fairly adequate for individual measurement."

Chapter VIII. There is sufficient correspondence between scores in the Berkshire Word Test and school marks to enable one to detect exceptional cases readily. The fact that this, like other tests reported, yields only a low correlation with school marks, does not destroy utterly its significance to the educator whose problem is so complex that no short test can measure all traits involved.

By reference to the age and grade scale in Chapter IV and the general percentile scale in Chapter VII, it is possible to evaluate in terms of a larger group any individual or class average.

CHAPTER X

CONCLUSION

From the facts stated above it is evident that vocabulary tests distribute with a fair degree of accuracy the ages and grades found in the junior and senior high schools of Pittsfield, Massachusetts. During the past five years so many pupils have been tested that it is safe to infer that such a test would give a correct relative rating for any similar group of pupils.

The Berkshire Word Test can be quickly given, one list requiring only ten minutes. A key renders scoring objective, thus eliminating many clerical errors. Interpretative tables furnish a quick evaluation of test scores that is valuable for educator, employer, or clinical psychologist.

Since the corresponding lines in this test are of nearly equivalent difficulty, it is possible to combine the units in a variety of ways to form tests that are relatively simple or relatively difficult. Interchange among the test units would reduce greatly the chance of errors due to learning the test.

Through further applications of the test the relative difficulty of the several units will be more fully established, but in its present form the Berkshire Word Test, when carefully administered, is an efficient aid in classifying both individuals and groups.

Some facts observed in this research may be of more value in the construction of similar tests than is the test presented here:

1. The double response required in the Berkshire Word Test increased greatly its efficiency without causing any marked reduction in the average score.

2. Errors were most frequent in column one.

3. Correct answers were most frequent in column two.

4. In such a test the relative difficulty of a line may be altered by changing the arrangement of the response words.

5. A knowledge of the relative difficulty of stimulus and response words is equally important.

Careful study of the vocabulary tests and word lists that have been published should reduce greatly the errors and therefore the time required to construct such a test as this presented here. Vocabulary tests appear in batteries of many

sorts, those for measuring special aptitudes as well as measures of general intelligence, the prognostic as well as diagnostic.

Tests of this kind furnish means for suggestive but not exhaustive study of mental ability. The large coefficient of alienation developed when the Berkshire Word Test was compared with school marks shows that there are many factors which are not common to the vocabulary test and the school marks.

A vocabulary test needs constant revision. Words that were familiar in 1922 have been displaced by others that were then in the lower portion of the list. "Drabble" in the Terman list is seldom known by a child who has lived in a day of short skirts. Table V shows the amount of displacement which occurred among 100 words in this research. Doubtless much of this change is due to sampling, but a part of it was caused by change of usage.

With increased precision vocabulary tests will become correspondingly valuable in clinical work. But no objective test of individuals assembled in groups can furnish adequate data for a complete study of intelligence. All such measuring scales in the hands of competent examiners furnish useful, comparable, and remarkably reliable data; but these results should be evaluated only after comparison with other personality studies.

The correlation of the vocabulary with scores in general intelligence tests is of such an order as to prove that the vocabulary, while contributing to the measurement of intelligence, does not measure all that is measured by tests of the omnibus type. In any test of general intelligence a vocabulary should form an important part but should not be relied on to measure all of intelligence.

APPENDIX

The Berkshire Word Test
Examiner's Guide
Directions for administering the test

Any teacher who is accustomed to handling standard intelligence tests will be able to administer the Berkshire Word Test satisfactorily. The best preparation for giving the test is to take it and score one's own paper. It is very important that conditions be kept uniform in all classes tested and that the same methods be used by all examiners so that results may be fairly compared and norms established. The examiner should, therefore, say exactly what is directed. If some explanation is required to make clear the meaning of page one, it is permissible to give this. Without an understanding of that page success in the test is impossible. If, however, after a careful explanation, a pupil fails to grasp the meaning of the directions, that fact suggests that the test is too difficult. Let such pupils do what they think is right; but in no case answer questions or give help after the work on pages two, three or four has been commenced. Try to see that the pupils are entering their answers in the parentheses, but give no other suggestions.

Method I

For use in grades seven through ten

Have monitors pass out the papers, face up, one to each pupil. Then say to the pupils, *"We are going to try some tests today to see how many words you know. We are passing out the test papers now. Do not open your paper nor turn it over until you are told to do so. Begin to read the first page and fill in the blanks that you find there:*

Your name; how old you are; your birthday; what date it is today; what grade you are in, 7A, 7B, etc."

Allow time for this writing, then say, *"Now let us read together what is printed on page one."*

Examiner reads aloud the directions given on page one of the test paper trying to see that the pupils understand what they are supposed to do. When the reading is completed, say, *"Is there anybody who does not understand the first page?"*

When the pupils understand page one, say, "*You will have thirty minutes for this test. This will give you ten minutes for each page. Turn to page two and begin.*"

At the end of ten minutes say, "*Go to page three.*" At the end of twenty minutes say, "*Turn to page four.*" At the end of thirty minutes say, "*Stop. Turn to page one and pass in your papers, face up.*"

The test may be given in three parts if it is more convenient. When this method is employed, ten minutes should be allowed for each page. At the close of each test period the papers should be collected and passed out again for each succeeding exercise.

Method II

For use in Grades eleven and twelve

While monitors pass out the papers, face up, one to each student, say, "*We are going to try some tests today to see how many words you know. Do not open your paper nor turn it over until you are told to do so. Begin to read the first page and fill in the blanks there.*"

Allow time for this writing; then say, "*Now read the directions printed on page one.*"

After two or three minutes, say, "*Is there anybody who does not understand page one?*"

If all seem to understand the task, say, "*You will be allowed 30 minutes for the test. That gives you ten minutes for each page. Turn to page two and do what you are told to do.*"

After 30 minutes say, "*Stop. Turn to page one and pass in your papers.*"

FORM A

Key to Page 2	Key to Page 3	Key to Page 4	Key to Page 2	Key to Page 3	Key to Page 4
1. (1)	1. (4)	1. (2)	26. (2)	26. (4)	26. (2)
2. (2)	2. (1)	2. (1)	27. (1)	27. (3)	27. (3)
3. (4)	3. (3)	3. (3)	28. (2)	28. (2)	28. (2)
4. (1)	4. (2)	4. (1)	29. (4)	29. (4)	29. (1)
5. (2)	5. (1)	5. (4)	30. (2)	30. (1)	30. (2)
6. (1)	6. (2)	6. (4)	31. (3)	31. (3)	31. (4)
7. (1)	7. (3)	7. (1)	32. (3)	32. (2)	32. (4)
8. (2)	8. (4)	8. (3)	33. (4)	33. (1)	33. (4)
9. (1)	9. (3)	9. (1)	34. (2)	34. (4)	34. (3)
10. (4)	10. (2)	10. (4)	35. (3)	35. (3)	35. (1)
11. (3)	11. (1)	11. (2)	36. (4)	36. (4)	36. (3)
12. (2)	12. (4)	12. (2)	37. (1)	37. (2)	37. (3)
13. (1)	13. (3)	13. (2)	38. (2)	38. (1)	38. (3)
14. (3)	14. (2)	14. (2)	39. (4)	39. (3)	39. (1)
15. (2)	15. (1)	15. (3)	40. (4)	40. (3)	40. (4)
16. (3)	16. (1)	16. (4)	41. (4)	41. (1)	41. (1)
17. (3)	17. (3)	17. (2)	42. (1)	42. (2)	42. (4)
18. (4)	18. (2)	18. (3)	43. (1)	43. (4)	43. (1)
19. (1)	19. (4)	19. (1)	44. (2)	44. (4)	44. (4)
20. (4)	20. (4)	20. (3)	45. (2)	45. (3)	45. (2)
21. (1)	21. (2)	21. (1)	46. (3)	46. (4)	46. (4)
22. (3)	22. (3)	22. (4)	47. (3)	47. (2)	47. (2)
23. (3)	23. (1)	23. (1)	48. (4)	48. (1)	48. (3)
24. (2)	24. (2)	24. (2)	49. (1)	49. (2)	49. (3)
25. (4)	25. (1)	25. (1)	50. (4)	50. (3)	50. (3)

Score = number right.

BIBLIOGRAPHY

1. Arlitt, Ada H. and Hall, Margaret, "Intelligence Tests Versus Entrance Examinations as a Means of Predicting Success in College," *Journal of Applied Psychology*, Vol. VII (1923), pages 330-338.

2. Bonser, F. G., Turner, M.R. and Burch, L. H., "Vocabulary Tests as Measures of School Efficiency," *School and Society*, November 13, 1915, pages 713-718.

3. Brooks, F. D., "The Accuracy of Intelligence Quotients from Pairs of Group Tests in the Junior High School," *Journal of Educational Psychology*, Vol. 18 (1927), pages 173-186.

4. Caldwell, H.H., "Adult Tests of the Stanford Revision Applied to College Students," *Journal of Educational Psychology*, Vol. 10 (1919), pages 477-488.

5. Cannon, J.G., "An Enquiry into the Relative Values of the Inventive and Selective Forms of Group Tests of Mental Capacity," *The Australasian Journal of Psychology and Philosophy*, Vol. 4, No. 2 (1926), pages 141-149. (Psychological Laboratory, Teachers College, Melbourne).

6. Chapman, J. C., "The Unreliability of the Difference between Intelligence and Educational Ratings," *Journal of Educational Psychology*, Vol. 14, pages 103-108.

7. Colvin, S. S. and Allen, R. R., "Mental Tests and Linguistic Ability," *Journal of Educational Psychology*, Vol. 14 (1923), pages 1-20.

8. Doran, E. W., "A Study of Vocabularies," *Pedagogical Seminary*, Vol. 14 (1907), pages 401-438.

9. Foster, R. R. and Ruch, G. M., "On Corrections for Chance in Multiple Response Tests," *Journal of Educational Psychology*, Vol. 18 (1927), pages 48-51.

10. Franz, S. I., *Handbook of Mental Examination Methods*. The Macmillan Company, New York. 193 pages. Second edition 1919.

11. Freeman, Frank N., *Mental Tests. Their History, Principles and Applications*, Houghton Mifflin Company, Boston, 503 pages. 1926.

12. Gates, A. I., "The Unreliability of M.A. and I.Q. Based on Tests of General Ability," *Journal of Applied Psychology*, Vol. 7 (1923), pages 93-100.

13. Gerlach, Fred M., *Vocabulary Studies*: Studies in Education and Psychology, No. 1, 123 pages. Colorado College, 1917.

14. Geyer, D. L., "Reliability of Ratings by Group Intelligence Tests," *Journal of Educational Psychology*, Vol. 13 (1922), pages 43-49.

15. Haddock, Nellie P., "The Use of the Stanford Achievement Test to Indicate High School Success," *School and Society*, January 22, 1927, pages 114-116.

16. Haggerty, M. R., *Haggerty Reading Examination: Sigma 3*. World Book Company, Yonkers-on-Hudson, New York. 1922.

17. Henmon, V. A. C., "An Experimental Study of the Value of Word Study," *Journal of Educational Psychology*, Vol. 12 (1921), pages 98-102.

18. Holley, C. E., *The Holley Sentence Vocabulary Tests*, Decatur Test Publishing Company, Decatur. 1926.

19. Inglis, A., *The Inglis Tests of English Vocabulary*, Ginn and Company, Boston. 1924.

20. Kelly, T. I., Ruch, G. M. and Terman, L. M., *Stanford Achievement Test*, World Book Company, Yonkers-on-Hudson. 1923.

21. Kennon, L. H. V., *Tests of Literary Vocabulary for Teachers of English*, Teachers College Contributions to Education, No. 223, Columbia University. 1926.

22. Kirkpatrick, E. A., "A Vocabulary Test," *Popular Science Monthly*, Vol. 70 (1907), pages 157-164.

23. Neher, H. L., "Measuring the Vocabulary of High School Pupils," *School and Society*, Vol. 8 (1918), pages 355-359.

24. Otis, A. S., *Otis Classification Test*, World Book Company, Yonkers-on-Hudson, New York. 1923.

25. ———— *Otis Group Intelligence Scale*, World Book Company, Yonkers-on-Hudson, New York. 1921.

26. ———— *Otis Self-Administering Tests of Mental Ability*, World Book Company, Yonkers-on-Hudson, New York. 1922.

27. ———— "Some Logical Aspects of the Binet Scale," *Psychological Review*, Vol. 23 (1916), pages 129-152.

28. ———— *Statistical Method in Educational Measurement*, World Book Company, Yonkers-on-Hudson, New York. 337 pages. 1925.

29. ———— "The Method of Finding the Correspondence between Scores in Two Tests," *Journal of Educational Psychology*, Vol. 13 (1922), pages 529-545.

30. Pintner, R., "The Measurement of Progress in Language Ability, *Journal of Educational Psychology*, Vol. 9 (1918), pages 270-277.

31. Pintner, R. and Fitzgerald, F., "An Educational Survey Test," *Journal of Educational Psychology*, Vol. 11 (1920), pages 207-223.

32. Pintner, R. and Paterson, D. G., "A Measurement of the Language Ability of Deaf Children," *Psychological Review*, Vol. 23 (1916), pages 413-416.

33. Root, W. I., "Correlation of Group Tests," *Journal of Educational Psychology*, Vol. 13 (1922), pages 286-292.

34. Schwesinger, G. C., *The Social-Ethical Significance of Vocabulary*, Teachers College Contributions to Education, No. 11, Columbia University, New York. 1926.

35. Seago, D. W., *An Analysis of Language Factors in Intelligence Tests*, Mental Measurement Monographs, Serial Number 1, pages 36-45.

36. Terman, L. M., *Terman Group Test of Mental Ability*, World Book Company, Yonkers-on-Hudson, New York. 1920.

37. Terman, L. M., *The Intelligence of School Children. How Children Differ in Ability. The Use of Mental Tests in School Grading, and the Proper Education of Exceptional Children*, Houghton Mifflin Company, Boston, 317 pages. 1919.

38. ———— *The Measurement of Intelligence. An Explanation of and a Complete Guide for the Use of the Binet-Simon Intelligence Scale.* Houghton Mifflin Company, Boston, 362 pages. 1919.

39. ———— Anderson, M., Chamberlain, M. B., Henry, M. B., and Kohs, S. C., "The Vocabulary Test as a Measure of Intelligence," *Journal of Educational Psychology*, Vol. 9 (1918), pages 452-466.

40. ———— and Childs, S. G., "A Tentative Revision and Extension of the Binet-Simon Measuring Scale of Intelligence," *Journal of Educational Psychology*, Vol. 3 (1912), pages 198-208, 277-289.

41. Thorndike, E. L., "Tests of Intelligence, Reliability, Significance, Susceptibility to Special Training and Adaptation to General Nature of the Task," *School and Society*, Vol. 9 (1919), pages 189-195.

42. ———— *The Measurement of Intelligence*, Bureau of Publications, Teachers College, Columbia University, New York, 616 pages. 1926.

43. ———— *The Teacher's Word Book*, Teachers College, Columbia University, New York, 134 pages. 1921.

44. ————— *Vocabularies of School Pupils*, New York Society; Contributions to Education, Vol. 1 (1924), World Book Company, Yonkers-on-Hudson, New York. 1924.

45. ————— "Word Knowledge in the Elementary School," *Teachers College Record*, Vol. 24 (1923), pages 438-445.

46. ————— and Symonds, P. M., "Difficulty, Reliability and Grade Achievements in a Test of English Vocabulary," *Teachers College Record*, (1923), Vol. XXIV, pages 438-445.

47. Trabue, M. R. and Stockbridge, F. P., *Measure Your Mind: The Mentimeter and How to Use It*, Doubleday, Page and Company, New York, 349 pages. 1920.

48. Weeks, A. L., *Mental Survey, Pittsfield, Massachusetts*, Annual Report of the School Committee, pages 41-63, 1922.

49. ————— "Terman Vocabulary as a Group Test," *Journal of Educational Psychology*, Vol. 12, pages 532-536.

50. Whipple, G. M., *Manual of Mental and Physical Tests*, part II, pages 274-283, 308-317, Warwick and York, Baltimore, 701 pages. 1915.

51. Woodworth, R. S. and Wells, F. L., *Association Tests*, Psychological Monographs, Vol. 13, No. 57, page 85. 1911.

52. Yerkes, R. M., Haggerty, M. R., Terman, L. M., Thorndike, E. L. and Whipple, G. M., *National Intelligence Tests*, World Book Company, Yonkers-on-Hudson, New York, 1921.

Addenda

53. Averill, L. A. and Mueller, A. D., "Some Results of Intelligence and Achievement Tests in English Given to Entering Normal School Students," *Educational Administration and Supervision*, January, 1927, pages 20-25.

54. Ohlson, David, "School Marks Versus Intelligence Rating," *Educational Administration and Supervision*, February, 1927, Vol. 13, pages 90-102.

55. Ruch, G. M. and Stoddard, G. D., *Tests and Measurements in High School Instruction*, World Book Company, Yonkers-on-Hudson, New York, 381 pages. 1927.

56. Wells, F. L., *Mental Tests in Clinical Practice*, World Book Company, Yonkers-on-Hudson, New York, 1927.

1799

1799